

从视频走向城市3D重构

A.阿克巴尔扎德⁺, J.-M. Frahm⁺, P. Mordohai⁺, B. Clipp⁺, C. Engels⁺, 盖洛普 (D. Gallup)⁺,
P.Merrell⁺, M. 菲尔普斯⁺, S.Sinha⁺, B.Talton⁺, L.Wang⁺, 杨⁺, 斯图文纽斯⁺,
杨扬⁺, G. Welch⁺, H. Towles⁺, D. Nist⁺ er⁺和M. Pollefeys⁺
⁺ 计算机科学系 ⁺ 计算机科学系

北卡罗来纳大学教堂山可视化与虚拟环境中心
肯塔基大学，美国列克星敦

美国教堂山

抽象

本文介绍了一种数据收集系统和处理管道，可通过视频对地理场景进行自动地理注册3D重建。该系统收集多个视频流以及GPS和INS测量，以便将重建的模型放置在地理注册的坐标中。除了在几何形状和外观方面均具有高质量外，我们还致力于实时性能。尽管目前我们的处理流程还远远不是实时的，但我们还是选择了一些技术并设计了可以在多个CPU和GPU上实现快速性能的处理模块，旨在在不久的将来实现实时性能。我们提出了设计系统的主要考虑因素以及处理流程的步骤。我们在系统捕获的真实视频序列上显示结果。

是一种自动系统，用于处理以无限制方式获取的大量视频数据。这迫使我们从实验室以外的视频中塑造出来，并实现一个可靠的系统。

视频采集系统由安装在车辆上的八个摄像头组成，每侧有四个摄像头。相机的分辨率为 1024 × 768

像素和30 Hz的帧频。每个四元组由指向侧面（垂直于行驶方向），对角线向前，向后和向上，最小重叠的摄像机组成，以实现较大的水平和垂直视野。另外，采集系统采用惯性导航系统（INS）和GPS来实现摄像机的地理定位。从我们的系统中进行地面重建的示例可以在图5和6中看到。1和2。

1引言

通常，城市的详细3D模型是由航拍数据制成的，其形式为距离或被动图像以及其他模式，例如来自全球定位系统（GPS）的测量。尽管这些模型可能对导航有用，但与地图相比，它们在可视化方面几乎没有提供其他信息。建筑物和其他地标不容易辨认，因为由于视角不好，这些立面无法从航空影像中重建。为了实现高质量的地面可视化，需要从地面捕获数据。一个自动生成纹理映射的地面3D模型的系统应该能够捕获大量数据，同时在街道上行驶并有效地处理这些数据。



图1.密集重建的示例。

在本文中，我们介绍了一种通过多相机系统捕获的数小时视频数据对城市场景进行全自动3D重建的方法。目标

整个采集系统包装在密封的容器中，该容器安装在车辆的后部。当车辆在城市环境中行驶时，



图2.一个城市街区的密集重建。

视频存储在Pod中的磁盘驱动器上。捕获会话后，驱动器从车辆的吊舱移至10 PC（双处理器）计算机集群以进行处理。我们的性能目标是在相同的时间处理多达6个小时的采集数据。

处理包括以下步骤：稀疏重建，在该稀疏重建中，根据视频和INS / GPS数据估算摄像机的地理定位姿态；密集重建，在此期间，根据视频数据和稀疏步骤的结果，计算出具有纹理映射的3D城市场景模型。

在稀疏重建中，使用运动技术的结构从视频数据估计摄像机的轨迹。目标是实现精确的相机姿势，以支持时间多视图立体，同时保持全局一致的地理记录轨迹无漂移。为此，对INS / GPS数据进行后处理，以获取经过过滤的车辆精确轨迹，这被称为平滑最佳估计轨迹（SBET）。SBET坐标系的原点和相机坐标系之间的SBET和手眼校准可提供对相机轨迹的可靠估计。

在密集的重建中，建筑物，地面和其他结构的表面是使用多视图立体技术估算的。此步骤的目标是即使在存在模糊或几乎没有表面纹理，遮挡或镜面反射的情况下，也要尽可能提供准确的表面。重建步骤分为多视图立体，该立体从多个视图生成深度图，其中单

gle参考视图和深度图融合，可解决多个深度图之间的冲突，并得出一致的表面描述。密集的重建阶段还使用视频输入为表面提供纹理。

在本文的其余部分安排如下。第1.1节讨论了相关工作。在第2节中详细描述了处理流程，在第3节中概述了具有INS / GPS记录功能的多摄像机捕获系统的不同系统方面，在第4节中回顾了实验结果，并在第5节中得出了结论。

1.1 以前的工作

研究社区已投入大量精力来使用传感器和模式的组合对人造环境进行建模。在这里，我们简略地回顾了基于地面成像的工作，因为它与我们的项目关系更密切。航空影像的工作量相等甚至更大。典型的目的是精确重建城市或考古遗址，包括几何和纹理，以便获得可用于可视化的模型，以大或小规模的测量形式进行定量分析，并有可能用于研究其演化通过时间。

满足对几何形状和外观建模的要求的自然选择是组合使用主动范围扫描仪和数码相机。Stamos和Allen [1]使用了这样的组合，同时还解决了注册这两种模式，将数据分段和拟合平面到点云的问题。El-Hakim等。[2]提出了一种在距离扫描仪，地面和航空图像以及CAD模型中选择最合适的方式的方法。”

uh和Zakhor [3]开发了一个与我们的方法非常相似，因为它也安装在车辆上，并且以连续模式捕获大量数据，这与以前的捕获场景的一些孤立图像相反。他们的系统由两台激光扫描仪组成，一台用于地图构建和注册，一台用于几何图形重建，一台数码相机用于纹理采集。Biber等人提出了一种结构相似但体积较小的系统，该系统也以连续模式运行。[4]。关于大规模城市建模的其他工作包括由Schindler等人进行的4D Atlanta项目。[5]，它还研究了模型随时间的演变。Cornelis等。[6]还开发了一种专门用于从安装在行驶中的车辆上的立体钻机重建立面的系统。

激光扫描仪具有直接提供准确的3D测量的优势。另一方面，它们可能笨重且昂贵。照相光度法和计算机视觉的几位研究人员解决了仅依靠无源传感器（相机）进行重建的问题，以提高系统的灵活性，同时降低了成本。

减小其尺寸，重量和成本。面临的挑战主要归因于2D测量中3D重构中文件记载的不准确性。为了获得有用的模型，可能必须与系统交互或进行简化的假设。第一次此类尝试是MIT City Scanning项目，其概述可在[7]中找到。Debevec等人提出了一种半自动方法，在该方法下将简单的几何图元拟合到数据中。[8]。即使没有对细节进行建模，而是将其视为纹理，也可以重建引人注目的模型。Rother和Carlsson [9]表明，在已知的固定平面在所有图像中都可见的情况下，多视图重建可以表述为线性估计问题。这种方法还需要手动操作。迪克等。[10]提出了一种自动方法，该方法从稀疏特征推断出分段平面，并考虑了诸如正交性和垂直性之类的约束。作者后来提出了一种更为复杂的基于MCMC的方法[11]，该方法将生成模型用于建筑物。它也是全自动的，但是受现有模型的限制，只能在少量图像上操作，通常是两到六个。[5]也采用了类似的高级推理。Werner和Zisserman [12]提出了一种自动方法，受[8]的启发，该方法通过检查经过改进的空间扫描算法[13]来支持平面和多面体在稀疏重建的基元上。作者后来提出了一种更为复杂的基于MCMC的方法[11]，该方法将生成模型用于建筑物。它也是全自动的，但是受现有模型的限制，只能在少量图像上操作，通常是两到六个。[5]也采用了类似的高级推理。Werner和Zisserman [12]提出了一种自动方法，受[8]的启发，该方法通过检查经过改进的空间扫描算法[13]来支持平面和多面体在稀疏重建的基元上。作者后来提出了一种更为复杂的基于MCMC的方法[11]，该方法将生成模型用于建筑物。它也是全自动的，但是受现有模型的限制，只能在少量图像上操作，通常是两到六个。[5]也采用了类似的高级推理。Werner和Zisserman [12]提出了一种自动方法，受[8]的启发，该方法通过检查经过改进的空间扫描算法[13]来支持平面和多面体在稀疏重建的基元上。

在过去的20年中，我们仅依靠无源传感器来解决该问题，它基于对计算机视觉界中运动和形状重构的结构的研究的经验。由于该文献太大，无法在此处进行调查，因此请参考感兴趣的读者[14，15]。我们项目的重点是开发一种全自动系统，该系统能够以连续模式运行，而无需从选定的角度捕获数据，因为捕获是从受限于城市街道有利位置的移动车辆上进行的。我们的系统设计还受到性能目标的驱动，即能够在等于采集时间的时间内对大型视频数据集进行后处理。

2加工管道

在下文中，我们将更详细地描述系统中使用的不同技术。处理流水线首先为视频的每一帧估计一个地理注册的摄像机姿势。我们通过确定连续视频帧中的2D-2D点对应关系来解决这一问题。然后，我们使用内部的相对相机几何

校准后的相机以为相机建立一个欧几里得空间。INS / GPS信息用于计算相机在地理空间坐标系中的位置。

计算完相机姿势后，我们将它们与视频帧一起使用以对输入图像执行立体匹配。这导致每个帧的深度图。这些深度图随后被融合以增强它们之间的一致性。处理流程的流程图如图3所示。

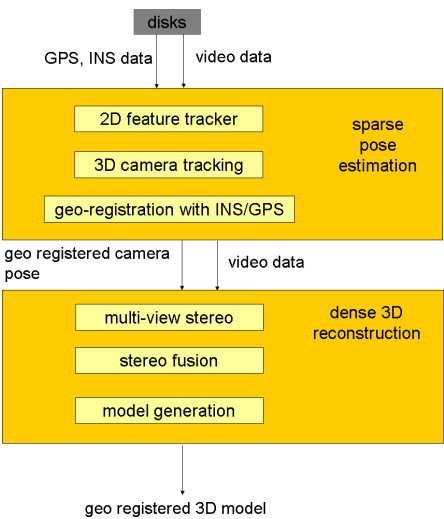


图3. 3D处理管道

2.1 2D特征跟踪

为了在连续视频帧之间建立2D特征对应关系，我们使用分层KLT跟踪器[22]来跟踪特征。为了实现视频帧率的实时跟踪，我们在GPU上使用了分层KLT跟踪器的实现[23]。平均需要30ms来跟踪一个目标中的1000个特征点 1024 × 768 ATI X1900图形卡上的图像。

跟踪技术的弱点是视差范围大，因为在相应的金字塔等级上，运动的流量假设少于一个像素会限制可捕获的运动量。因此，视差较大的视频帧给KLT跟踪器带来了问题。因此，我们还可以使用类似于[24]的检测和匹配跟踪器。它的优势在于，它可以非常快速地搜索较大的视差范围，其速度比从磁盘获取视频的速度还快。它的缺点是在嘈杂的低纹理条件下，检测的可重复性并不总是可靠的（[25]现象类似）。

2.2 3D摄像机跟踪

我们正在研究和开发几种方法，根据INS / GPS数据的可用性从2D特征轨迹确定相机姿态。我们希望在没有此类数据的情况下我们的系统能够正常运行。

当INS / GPS数据不可用时，我们将按照[18]的方法使用仅视觉摄像头跟踪算法。Briefly说过，我们可以使用三个视图的相对姿势来初始化摄像机跟踪器，并给出三个视图中的功能对应关系。使用计算出的相机姿势对这些对应关系进行三角测量。使用RANSAC和假设生成，使用从2D特征到3D世界点对应关系的约束来计算其他姿势。当新的世界点可用时，将使用新视图重新划分它们。

为避免累积的漂移，系统会定期使用一组新的三个视图重新初始化系统。我们利用一台重叠摄像机的约束将新姿势缝合到旧坐标系中。剩余的自由度是新旧坐标系的比例。使用两个坐标系中的相应三角点进行估算。

所有姿态估计方法都使用抢占式RANSAC和局部迭代修正以提高鲁棒性[26]。实际上，除非我们使用束调整来调整姿势，否则系统必须经常重新初始化。使用束调整时，姿势估计对测量噪声不太敏感，从而导致更少的重新初始化。

2.3使用INS / GPS数据进行地理注册

为了确定3D模型中要素的地理注册坐标，我们使用INS / GPS数据。INS / GPS测量系统配有GPS接收器，陀螺仪和加速度计。它可以对安装了摄像头的车辆的位置和方向进行高度精确的测量。

欧几里德变换（将被称为手眼校准）将地理位置系统的中心映射到每个摄像机的光学中心。最初，每个摄像机都保留自己的坐标系。每个相机的第一帧的光学中心是原点，并且光轴和第一像平面的轴用作轴。通过将视频序列中第一和第二摄像机位置之间的距离设置为单位长度可以任意选择比例。

我们的地理注册的第一个实现方法是计算视觉坐标系统中给定摄像机的姿态与世界上每个摄像机的姿态之间的相似度转换（旋转，平移和比例）。这种方法在处理基于视觉的相机姿势估计中的漂移方面存在困难，因为它仅限于一种刚性

所有帧的转换。

目前，我们正在开发第二种方法，该方法通过融合地理位置测量和使用Kalman滤波器或通过束调整来跟踪2D特征来克服这些限制。这些方法有望胜过相似度转换，地理配准技术。

2.4多视角立体声

多视图立体声模块将来自单个视频流的相机姿势和图像作为输入，并为每个帧生成深度图。它使用Collins [13]的平面扫描算法，这是一种有效的多图像匹配技术。从概念上讲，飞机会沿着预定方向（通常与光轴平行）逐步扫过空间。在平面的每个位置，所有视图都投影在该平面上。如果平面上的某个点位于正确的深度，则根据亮度恒定性假设，投射到该点上的所有像素都应具有恒定的强度。我们通过对参考图像中定义的平方聚合窗口中的绝对强度差求和，来测量这种一致性，并为此计算深度图。



图4.渲染为3D模型的立体深度图。(a)仅正面平行扫描。(b)多个扫描方向。

传统的平面扫描技术通常仅扫视正面平行的平面，这不能解释在非正面平行的表面中观察到的透视。我们通过在多个方向扫描平面来扩展算法，这些方向的方向与我们希望在场景中观察到的平面（例如地面和建筑立面）对齐。我们可以通过假定车辆平行于地面和立面，并且立面是垂直的并以直角相交，来预先推断出地面和立面的方向。图4说明了从多个方向进行扫描所获得的改进。除了可以同时处理多个图像的能力之外，平面扫描立体算法还可以轻松实现

移植到GPU以获得非常快的性能[27]。

2.5 立体声融合

多视图立体声为每个参考帧提供了深度图。由于我们在立体声过程中不强制深度图之间保持一致，因此我们需要在单独的立体声融合步骤中强制执行。融合有两个目的：通过确保同一点的估计与多个深度图一致来提高深度估计的质量，并通过将多个深度估计合并为一个来产生更经济的表示。相关工作包括体积[28, 29]，基于补丁的[30]，基于视点的[31]和基于像素的[32]方法。我们选择一种基于观点的方法，该方法受到中值融合算法的启发[17]。体积方法是不切实际的，因为要满足我们的图像分辨率和精度要求，将需要大量的体素。取而代之的是，我们对参考视图的每个像素使用一组深度假设。将图像平面作为参考很有用，因为这样可以通过对相邻像素的深度估计进行三角剖分来快速生成最终网格。

给定来自一组连续帧的深度图，立体声融合步骤可解决计算深度之间的冲突，并生成深度图，其中大部分噪声已被消除。融合非常有效，因为在大量帧中可见点，从而为每个点提供了多个深度估计。即使每个深度估计都是通过简单，快速的立体声算法生成的，但这些估计之间的共识通常还是非常准确的。融合旨在解决如图5所示的冲突。在图5.a中，深度估计 X 在参考相机的视野中 C_{ref} 遮挡深度图 d_{ref} 相机的 C_i ，而在图5.b中，深度估计 X 参考摄影机的摄影机通过摄影机的深度图 P_{ref} 。两种情况都是不可能的，应予以纠正。

一种方法是[17]的中值融合算法。输入是一组深度图，其中一个（通常是中间的一个）用作参考。该算法旨在根据每个像素选择最佳深度估计，稳定性，这取决于每个深度假设与其他深度图之间冲突的数量和类型。对于参考深度图的每个深度估计，它会遮挡的其他深度图的数量（图

5.a）或通过（图5.b）。相对于参考深度图，对所有其他深度图的每个深度估计重复此过程。所选（最稳定）的解决方案定义为最小深度，对于该最小深度，至少有多少个深度图会阻塞它，而通过的深度图就越多。

我们也在研究一种类似的方法，该方法考虑了每个深度估计的可信度。立体声可以在部分图像中产生更准确的深度估计

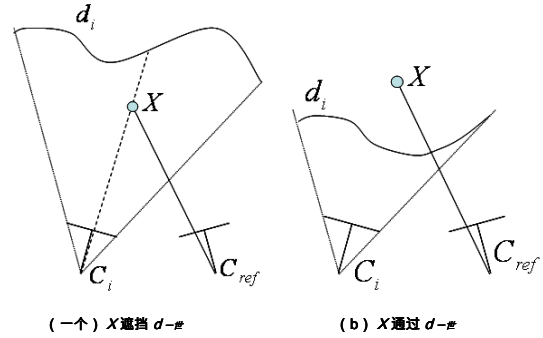


图5.深度图之间的冲突插图，必须通过融合来解决

有更多纹理且没有遮挡的年龄。通过在平面扫描阶段计算出的成本的最小值，可以明确表示匹配是否存在。另一方面，模糊匹配的成本函数具有多个最小值。基于成本函数的形式，我们可以为每个深度估计分配一个置信度值。在融合过程中，根据其他深度估计是支持还是矛盾，此置信度会增加或减少。低置信度的深度估计将替换为与置信度更高的邻居更紧密地一致的估计。这种方法会导致偏向于光滑的表面，这可能是合理的，因为我们强调的是重建建筑物，并且在大多数深度估计错误的区域（例如均匀的墙壁和街道）非常有效。另一方面，平滑度是以细化细节为代价的，而置信度的计算则需要额外的处理周期。我们在立体融合方面的未来工作将集中在更快地实现我们的算法和更好的方法来解决使用多个视图的遮挡问题上。

根据融合的立体深度图，我们在地理配准坐标中生成了一个三角的，纹理映射的3D模型。

3 视频采集系统

车载视频采集系统由两个主要子系统组成-8台摄像机数字录像机（DVR）和Applanix INS / GPS（POS LV型）导航系统。DVR将原始图像流式传输到磁盘，并且Applanix系统跟踪位置和方向，因此可以在通用的地理注册坐标系中创建后处理中生成的3D模型。

DVR由八个Point Grey Research（PGR）跳蚤彩色摄像机构成，在车辆的每一侧都有一个四倍的摄像机，如图6所示。

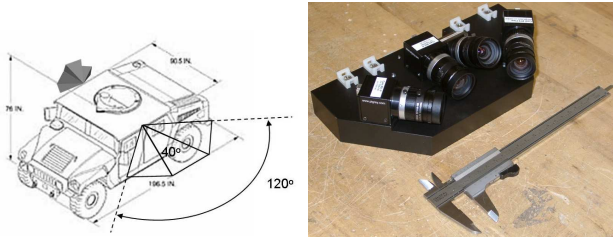


图6. (a) 悍马车上的摄像头四倍位置 (b) 摄像头四倍的特写镜头。

视野大约 $40^\circ \times 30^\circ$ ，在四个摄像头内，摄像机的视场重叠最小。如图6所示，三个摄像机安装在同一平面上，以创建大约120度的水平FOV。第四个摄像机向上倾斜，以使侧面摄像机产生大约60度的整体垂直FOV。

八个IEEE-1394摄像机连接到Concurrent Technologies的八个基于Windows-Intel VME的处理器模块，每个模块都具有大容量SATA数据驱动器。八摄像头DVR能够流式传输到磁盘 1024×768 30Hz的拜耳模式图像。使用PGR的IEEE-1394同步单元同步所有摄像机上的CCD曝光。

记录每个视频帧后，一个DVR处理器模块还会记录来自Applanix导航系统的GPS时间戳消息。这些消息事件也可以通过其中一台摄像机输出的外部TTL信号与CCD曝光同步。在后处理中，此GPS时间戳与后处理的INS / GPS数据相关联，以提供导航系统和每个摄像机的位置和方向的平滑最佳估计。当然，后者需要了解相机和在系统校准过程中建立的INS / GPS系统之间的手眼校准。

相机校准 相互校准摄像机是理想的，以融合由车辆上的八个摄像机的独立流创建的模型。此外，为了对从捕获的视频重建的所有3D模型进行地理配准，需要相对于INS坐标系进行校准。

当前，包括镜头畸变估计在内的照相机内在函数是使用平面棋盘和Ilie和Welch [33]基于OpenCV库构建的工具包来计算的。该解决方案提供了一种自动化程度更高的校准程序，需要最少的用户交互。

相机外部关系的评估由于以下事实而变得很复杂：

四个摄像机的视觉视野。目前正在采用的方法是将相机放置在3D特征点的较大范围内。3D特征点实际上是由一系列投影仪创建的，这些投影仪用棋盘图案照亮了摄像机前面的3D显示表面。每个特征点的3D位置是使用一对单独校准的立体摄像机确定的。

相对于INS，每个摄像机四倍的手眼（外部）校准必须同时包含摄像机四倍和车辆上安装的导航系统。最初，我们计划在经纬仪中确定相对位置的现实世界中使用经过校准的特征点。通过从众多车辆位置查看特征点，这些位置的位置和方向由Applanix系统估算，然后我们将能够建立每个摄像机相对于INS坐标系的旋转和平移关系。无疑将需要捆绑调整方法。未来的研究将集中于消除使用预先调查的特征点的限制。

4结果

图7至图10说明了当前正在重建的3D模型的保真度。除对图10中的一个进行地理注册外，所有模型的计算都没有任何INS或GPS信息。用于这些重构的典型设置如下：用于跟踪的特征数为500，用于每个深度图计算的图像数为11，融合的深度图数为13。



图7.前摄像机的密集重建。



图8.侧面摄像头的密集重建。



图9.侧面摄像头的密集重建。



图10.侧面摄像头的地理注册密集重构。

5总结与结论

我们已经描述了一种系统，该系统旨在通过多摄像机系统和INS / GPS测量捕获的视频来进行实时，密集，地理注册的3D城市重建。无论是否安装INS / GPS传感器，初始重建结果的质量都是非常希望的。未来的挑战包括提高地理配准的准确性，改善相机轨迹的束调整或卡尔曼滤波的过程，在多个视频流之间配准和融合重建图像，通过将操作移植到GPU来加快处理速度以及增强处理管道，使其更强大。潜在的长期方向是变更检测以及使用在不同日期获取的视频执行增量模型更新的功能。

致谢 这项工作得到了DARPA在UrbanScape项目的支持，该项目由上汽集团地理空间技术信息部牵头。

参考文献

- [1] I. Stamos和PK Allen, “几何和纹理重新掩盖了大规模的场景,” *计算机视觉和图像理解*, 卷 88号 2, 第94–118页, 2002年。
- [2] SF El-Hakim, J.-A. 贝拉拉丁, 皮卡德 (M. Picard) 和兽医 (A. Vet) - 撕毁“遗产遗址的有效3D建模”, 在 *第四届3D成像与建模国际会议*, 2003, 第302–309页。
- [3] C.Früh和A. Zakhor, “一种大规模的，基于地面的城市模型获取的自动化方法”, *诠释计算机视觉杂志*, 卷 60号 1, 第5-24页, 2004年。[4] P. Biber, S. Fleck, D. Stanek, M. Wand和W. Strasser, “在室内和室外环境中灵活地获取3d模型的移动平台的首次体验-waegel”, *ISPRS V / 4工作组会议记录: 3D-ARCH 2005年2005*。
- [5] G. Schindler, P. Krishnamurthy和F. Dellaert, 《城市环境中运动产生的基于线的结构》, *3DPVT*, 2006年。
- [6] N. Cornelis, K. Cornelis和L. Van Gool, “快速com-导航城市的可视化城市模型,” *诠释 Conf. 关于计算机视觉和模式识别*, 2006年。
- [7] S. Teller, “自动城市模型获取: 项目的原理和地位,” *图像理解研讨会*, 1998年, 第455-462页。

- [8] P. Debevec, C.J Taylor和J. Malik, “建模照片中的渲染和渲染架构：基于几何和图像的混合方法,” *SIGGRAPH*, 1996年, 第11-20页。
- [9] C. Rother和S. Carlsson, “线性多视图重新使用参考平面进行构造和摄像头恢复,” *诠释 计算机视觉杂志*, 卷 49 号 2-3, 第117–141页, 2002年。
- [10] A.R. Dick, P.H.S. Torr和R. Cipolla, “自动3d建筑建模”, 在 *英国机器视觉会议*, 2000年, 第273-289页。
- [11] A.R. Dick, P.H.S. Torr和R. Cipolla, “建模以及来自多个图像的建筑诠释,” *诠释 计算机视觉杂志*, 卷 60号 2, 第111–134页, 2004年。
- [12] T. Werner和A. Zisserman, “新技术从照片中自动进行建筑重建”, 在 *欧洲会议 在计算机视觉上* 2002年, 第II页: 541-555。
- [13] R.T. Collins, “真正的多图片匹配,” 中 *诠释 Conf. 关于计算机视觉和模式识别*, 1996年, 第358-363页。
- [14] R. Hartley和A. Zisserman, *多视图几何 尝试计算机视觉*, 剑桥大学出版社, 2000。
- [15] O.D. Faugeras, *三维计算机视觉：几何观点* 麻省理工学院出版社, 1993年。[16] M. Pollefeys, R. Koch和L. Van Gool, “自-校准和度量重建, 尽管摄像机的固有参数变化不定且未知,” *诠释 计算机视觉杂志*, 卷 32号 1, 第7-25页, 1999年。[17] D. Nistér, “从视频序列中估计相机运动, 根据未校准的视频序列自动进行密集重构”, 博士学位论文, 皇家技术学院, 斯德哥尔摩, 瑞典, 2001年。[18] D. Nistér, “有效解决五点相对姿势问题”, *IEEE Trans. 关于模式分析和机器智能*, 卷 26号 6, 第756–777页, 2004年。
- [19] D. Nistér, “从图像和视频自动被动恢复3d”, *3DPVT*, 2004年, 第438-445页。
- [20] M. Pollefeys, L. Van Gool, M. Vergauwen, F. Verbiest, K. Cornelis, J. Tops和R. Koch, “使用手持摄像机进行视觉建模” *诠释 计算机视觉杂志*, 卷 59, 没有 3, 第207–232页, 2004年。
- [21] M. Pollefeys, L. Van Gool, M. Vergauwen, K. Cornelis, F. Verbiest和J. Tops, “基于图像的3d记录考古领域的工作,” *计算机图形学与应用* 卷 23号 3, 第20–27页, 2003。
- [22] B.D. Lucas和T. Kanade, “迭代图像注册-立体视觉中的定位技术”, *诠释 联合会 关于人工智能*, 1981年。
- [23] S. Sinha, J.-M. Frahm和M. Pollefeys, “基于GPU视频功能跟踪和匹配”, 技术. Rep. TR06-012, 北卡罗来纳大学教堂山分校, 2006年5月。[24] D. Nistér, “视觉地面车辆应用的里程表,” *Journal of Field Robotics*, 卷 23号 2006年 1月1日。[25] D. Lowe, “来自比例尺的独特图像功能-不变的关键点,” *诠释 计算机视觉杂志*, 卷 60号 2, 第91-110页, 2004年。
- [26] D. Nistér, “用于实时结构和运动估计的抢先RANSAC”, 在 *诠释 Conf. 在计算机视觉上* 2003, 第一卷 1, 第199–206页。
- [27] R. Yang和M. Pollefeys, “多分辨率实时商品图形硬件上的立体声,” 中 *诠释 Conf. 关于计算机视觉和模式识别*, 2003年, 第I页: 211-217。
- [28] B. Curless和M. Levoy, “从范围图像构建复杂的模型,” *SIGGRAPH*, 卷 30, 第303-312页, 1996年。[29] M.D. Wheeler, Y. Sato和K. Ikeuchi, “共识”用于从多个范围图像对3d对象建模的曲面,” 中 *诠释 Conf. 在计算机视觉上* 1998年, 第917–924页。
- [30] P.V. Fua, “从多个立体视图到多个3-d表面” *诠释 计算机视觉杂志*, 卷 24号 1, 第19-35页, 1997年。
- [31] P.J. Narayanan, P.W. Rander和T. Kanade, “使用密集立体声构建虚拟世界”, 在 *诠释 Conf. 在计算机视觉上* 1998年, 第3-10页。
- [32] R. Koch, M. Pollefeys和L. Van Gool, “多视角-来自未校准视频序列的点立体声,” *欧洲会议 在计算机视觉上* 1998年, 第I页: 55–71。
- [33] A. Ilie和G. Welch, “确保色彩一致性跨多个摄像机” 中 *ICCV*, 2005年, 第1268页–1275。