中国科学院大学研究生课程讲义

课程： 计算机视觉

（第 4 章： 基于图像的三维重建）

教 研 室：脑科学与智能技术

教　　师：　　胡占义

编写时间：　　2019-8-22

第 4 章： 基于图像的三维重建

摘要：这是国科大春季学期为研究生开设的《计算机视觉》课程讲义的第 4 章。本章主要介绍从多幅无序图像如何对场景进行三维重建。本章分为 4 个主要部分：首先对三维重建原理，特别是分层三维重建原理进行介绍；鉴于当前 5-点算法在三维重建中的重要性，接着对 5-点算法进行简介；在此基础上，分别对稀疏三维重建和稠密三维重建进行比较详细的介绍。本章内容覆盖了从二维图像重建三维场景的主体内容和代表算法。通过对本章内容的学习，读者会对 “基于图像的三维重建”有一个比较系统全面的了解。

尽管深度学习在计算机视觉领域取得了突破性进展，很多过去优秀的算法均已经被基于深度学习的方法取代，但在从多幅图像对场景的三维重建领域，目前基于深度学习的方法仍无法与基于多视几何的方法相媲美。笔者 2018 年曾在《China Science: Information Science》以《Position paper》形式撰文（Dong et al. 2018: **Learning Stratified 3D reconstruction**）指出，这主要是由于不管用什么方法，图像之间的特征误匹配是不可避免的，而目前深度学习方法无法内嵌诸如 RANSAC 类的误匹配剔除模块，从而导致任何图像特征间的误匹配均会传输到最终的输出结果中，从而导致重建错误。事实上，目前基于深度学习的视觉定位，也无法与基于几何的方法相媲美，其原因与基于图像的三维重建相同。从这个意义上说，深度学习不是万能的，笔者估计在基于图像的三维重建方面，在短期内深度学习仍很难与基于几何的方法相媲美。

几何方法的优势在于“机理清晰”，“可解释性强”，“重建精度高”，不足是“计算比较复杂”，无法处理“无纹理的区域”。

目录

什么是基于图像的三维重建

首先，在介绍主要内容之前，有必要对“基于图像的三维重建”（image based 3D reconstruction）和“基于图像的三维建模”（image based modeling）进行区别。一般来说，基于图像的三维重建仅仅在于恢复“图像成像过程中丢失的深度信息”，即输出的是“三维点云”，这些三维点云并没有上升到“物体”（object）的概念. 比如，对一个办公室场景进行三维重建，可以得到办公室对应的三维点云，但我们并不知道这些点云中那些点云对应茶几，椅子等物体。要获得这些“物体”信息，还需要进行后续的物体分割等。“基于图像的建模”文献中一般指“对一个特定物体进行三维重建”，而且一般指每幅图像对应的摄像机位置和姿态均已标定好。基于图像的三维重建主要是计算机视觉领域的一个术语，而基于图像的建模更多是计算机图形学领域的一个术语。目前基于图像的三维场景语义重建（Semantic 3D Scene Reconstruction），则意味着不仅要从图像恢复三维点云，而且至少每个三维点对应的物体类别信息同时需要恢复。在有些情况下，甚至要输出“物体的参数模型”。

如对一个长方体类物体，要从点云拟合出其“姿态信息和长、宽和高等形状信息”。所以，基于图像的场景建模要比基于图像的三维重建包含更多的内涵，希望大家不要产生混淆。因此，本章内容也仅仅是从图像恢复场景的三维点云。所谓“稠密重建”，是指重建的点云为稠密点云，即每个像素都有对应的一个三维点。“稀疏重建”是指重建的点云为稀疏点云，“稀疏”是一个比较模糊的概念，可以有不同的稀疏程度。

4.1： 分层三维重建的原理

分层三维重建（stratified 3D reconstruction）是计算机视觉领域一种重要的三维重建理论。尽管随着相机加工基础的进步，传统小孔成像模型（the pinhole camera model）下的 5 个相机内参数仅仅“焦距”一个内参数需要标定且该焦距在拍摄图像时的具体值可以从图像的头文件中读出，因此相机的内参数可以认为是已知的，此时，从二幅图像之间的对应点可以通过“5-点算法”分解出“二幅图像对应相机之间的 5 个外参数（旋转矩阵和平移向量）”，即从图像对应点可以通过“5-点算法”直接进行三维重建，而不再需要“分层”进行重建，但“分层三维重建”由于其理论的优美性和在计算机视觉发展历程中所起的作用，本章首先对其原理进行简单介绍。

“基于图像的三维重建”就是指从多组、多幅图像之间的对应点恢复其对应的三维空间点的过程。“分层三维重建”就是指这个“重建三维点”的过程要分步进行。如图 4.1 所以，在分层重建框架下，从图像对应点首先恢复空间点在射影坐标系下(projective space) 的坐标，然后将射影坐标提升（upgrade）到仿射空间(affine space)下的坐标，进而提升到度量空间（metric space）下的坐标。这里“提升”是指“仿射空间是射影空间的子空间”。同理，度量空间是仿射空间的子空间。这里度量重建是指“绝对尺度未知”的欧几里德重建。因为从理论上来说，仅仅从图像无法恢复空间物体的绝对尺度大小。如一个离相机远一点的球体与一个离相机近一点的球体可以成同样的像，所以仅仅从图像无法恢复物体的绝对尺度，所以，从理论上来说，从图像仅仅能够达到度量空间下的重建。

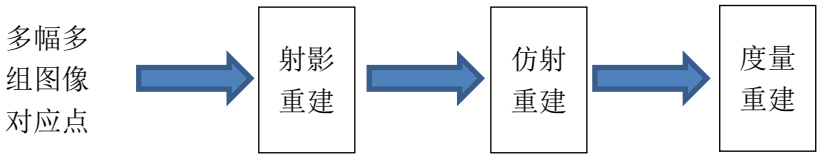

图 4.1： 分层三维重建

射影重建（Projective Reconstruction）

射影重建指从图像对应点重建的空间点的坐标与物体真正的三维坐标之间相差一个三维射影变换（三维射影变换矩阵就是一个一般的 4*4 的矩阵）。令给定物体坐标系下某个点的三维齐次坐标为：$\boldsymbol{X_i}$，重建点的其次坐标为 $\boldsymbol{Y_i}$，如果对所有的点 i，均有：$\boldsymbol{X_i} \sim \boldsymbol{A}\boldsymbol{Y_i}\ i = 1,2,\cdots,\mathrm{N}$，其中矩阵$\boldsymbol{A}$ 为一个未知（但唯一）的一般 4*4 的矩阵，则称 $\boldsymbol{Y_i}$ 为

X_i 的射影重建。注意，这里N 要足够大。很显然， 如果$N<6$，则矩阵A 不唯一。另外，符号“$\sim$” 表示相差一个常数因子下的相等。

仿射重建（Affine Reconstruction）

当上面矩阵A 退化为一个仿射矩阵时，此时称为仿射重建。

度量重建（Metric Reconstruction）

当上面矩阵A 退化为一个相似变换矩阵时，此时称为度量重建。注意，相似变换(similarity transform) 与刚体变换（rigid transform）不同。刚体变换前后物体的大小不变，而相似变换同时具有尺度的变化。文献中经常有人把刚体变换与相似变换不加区分，这是不严格的。

不同空间下重建结果的性质

射影空间下，直线之间的平行和垂直关系均不再保持。仿射空间下，能够保持直线之间的平行关系，但不能保持直线之间的垂直关系。度量空间下，既可以保持直线之间的平行关系，也可以保持直线之间的垂直关系。正像上面所说，此时仅仅是物体的绝对尺寸无法确定。

图 4.2 给出了一个长方体在不同重建下的示意图。

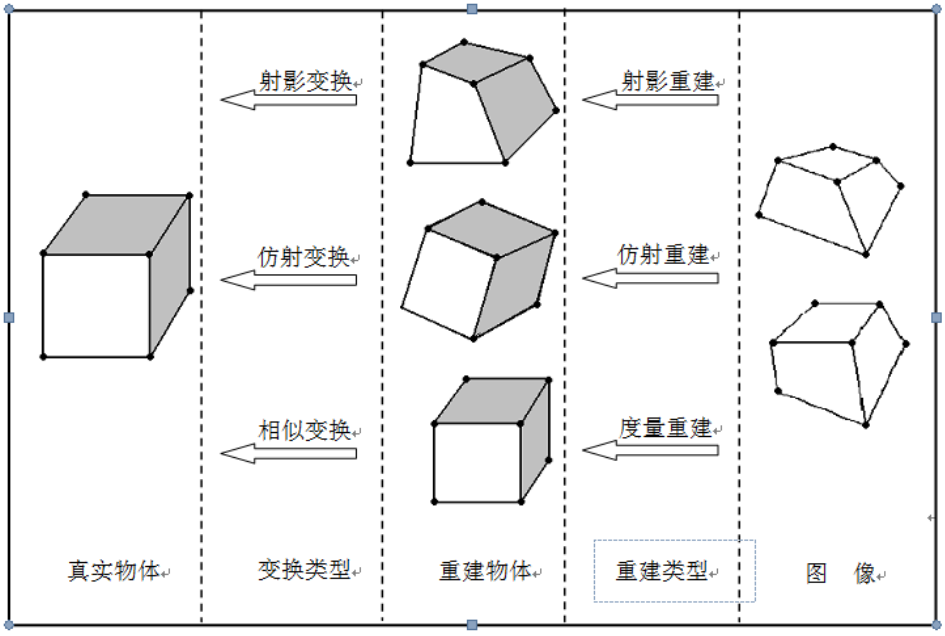

图 4.2： 不同重建下物体直线之间平行和垂直关系的保持特性

分层重建理论的优美性

基于图像的三维重建的目标在于从二维图像恢复“度量空间下的三维点云”，该重建过程最大的问题是鲁棒性差，从而很难在诸如机器人等要求高精度和可靠性的领域得到实际应用。这也是为什么马尔计算视觉从上世纪 90 年代初受到人们质疑和批评的缘故（见第一章：

计算机视觉简介）。 “分层”重建的好处在于从图像到度量空间重建的过程中，不是直接从图像到度量空间，而是分步进行：如图 4.1 所示，先从图像到射影空间，然后到仿射空间，最后到度量空间。 由于每一步重建过程中涉及的未知数个数相对少，所以重建过程的鲁棒性相对要高。

不同空间下的待重建量

分层重建有一套完整的理论，涉及“绝对二次曲线”（The absolute conic），“绝对二次曲面”（The absolute quadric）等三维射影空间的抽象概念，这里很难用较短的篇幅介绍清楚，感兴趣的学生可以参阅 Hartley 和 Zisserman 2000 年首次出版的关于多视几何的书 (Hartley & Zisserman 2000). 该书对多视几何理论，包括分层重建理论，进行了非常清晰的介绍。不管是摄像机自标定，还是分层重建，本质上依据的原理均是：绝对二次曲线（或绝对二次曲面）的像仅仅取决于相机内参数而与相机运动无关。 直观上说，由于绝对二次曲线（二次曲面）是无穷远平面上的量，相机运动是有限的，所以这些量在相机的投影不应该与相机运动有关。 当然，这是一种直观解释，射影几何有一套完整的对应理论，这里不再进一步介绍。

理论上说，从图像对应点进行射影重建，就是确定射影空间下每幅图像对应的 3*4 投影矩阵的过程，目前，文献中已有很多优秀算法。我个人觉得，射影重建已经是一个得到很好解决的问题，似乎已没有太多的研究价值。

仿射重建，在于确定无穷远平面在射影重建下（某个特定射影坐标系）对应坐标向量的过程（$(a \quad b \quad c \quad 1)^T$, 3 个未知元素）。无穷远平面在度量空间和仿射空间的坐标向量形式是已知的（即$(0 \quad 0 \quad 0 \quad 1)^T$），该向量在射影空间会随着所选取的射影坐标系的变化而变化，所以，该向量在射影重建下的形式会随着射影重建坐标系选取的不同而不同。事实上，所谓的仿射重建，就是要把射影空间对应的无穷远平面向量 $(a \quad b \quad c \quad 1)^T$ “拉回到” $(0 \quad 0 \quad 0 \quad 1)^T$ 的形式，此时需要确定(a,b,c)三个未知量。

仿射重建是分层重建最困难的一步。1998 年，Mark Pollefeys 等因提出所谓的模约束理论（modulus constraint），获得当年的马尔奖（Marr Prize）。模约束理论就是指无穷远平面在二幅图像之间的单应矩阵（infinite homography，M_∞）与一个旋转矩阵（$\mathbf{R}$）相似的事实，即 $\mathbf{M}_\infty \sim \mathbf{K}\mathbf{R}K^{-1}$. 基于此，Pollefeys 等给出了一个关于$(a,b,c)$ 未知量的三元四次,称之为模约束, 的约束方程。有兴趣的读者可参阅 Pollefeys 等的工作（Pollefeys et al. 1998）。

从仿射重建到度量重建，本质上在于确定摄像机的内参数矩阵，即摄像机的自标定过程。这里需要说明的是，一旦投影矩阵已知，当已知图像对应点后，对应的空间点可以很容易计算，所以，表 4.1 中的待重建量没有列出空间点。实际中，空间点是真正需要重建的对象，而确定投影矩阵，无穷远平面和内参数矩阵等，仅仅是为重建空间点服务。 下表 4.1 给出了不同重建下的待重建量：

表 4.1：不同重建下的待重建量

输入	射影重建	仿射重建	度量重建
多幅多组图像对应点	每幅图像对应的 3*4 投影矩阵。已有大量优秀算法	无穷远平面对应的 3-D 向量。这是分层重建最困难的一步	相机的内参数

正像本章开始所说的，尽管分层重建理论优美，在计算机视觉的发展中曾有过辉煌的经历，但由于相机制造工艺的提高，目前相机的焦距可以从图像头文件中读出，因此基于二幅图像之间的本质矩阵（Essential matrix）约束，通过“5-点算法”（5-point algorithm）就可以求解二幅图像之间的外参数（旋转和平移向量），直接进行三维重建，所以目前的重建方法已基本不再使用分层重建方法. 鉴于 5-点算法在基于图像的三维重建中的重要性，下面将对 5-点算法进行单独介绍。

4.2 5-点算法简介

5-点算法是指相机内参数已知的情况下，已知二幅图像之间的 5 组图像对应点，如何求取二幅图像之间的本质矩阵，进而分解出对应的旋转矩阵和平移向量的一种方法。5-点算法由 David Nister 于 2004 年提出（Nister 2004），现已成为为基于图像的三维重建的一种广泛使用的方法。5-点算法涉及大量的数学推导，有兴趣的学生可参阅这篇文章。下面仅仅给出基本原理.

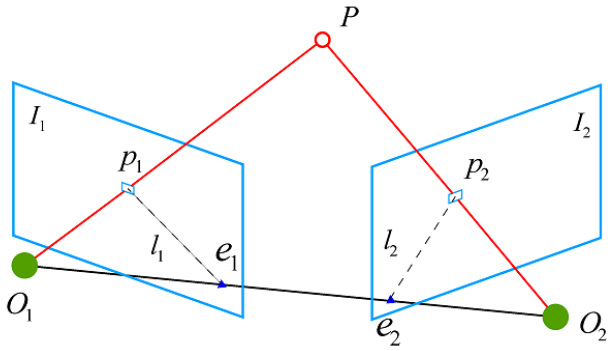

图 4.3： 对极几何图

如图 4.3 所示，$p_{1,i}$ 是空间点P_i在左图像上的投影，$p_{2,i}$ 是空间点P_i在右图像上的投影。由于假定相机的内参数已知，此时$p_{1,i},p_{2,i}$可以通过已知的内参数矩阵转化为归一化的对应齐次坐标：$p'_{i,j}=K_i^{-1}p_{i,j}\ \ i=1,2;j=1,2,\cdots N$，其中 K_1 和 K_2 为左右图像的内参数矩阵。假定左右图像之间的旋转矩阵和平移向量为R,T,则对应的本质矩阵为：$E=T\times R=[T]_\times R$ ，

其中$[T]_\times$表示由向量T构成的反对称矩阵（skew-symmetric matrix）。根据对极几何约束知：对所有的归一化对应点对$\{p'_{1,j} \sim p'_{2,j},\ j=1,2,\cdots,N\}$，下面的约束等式成立：

$$\left(p'_{1,j}\right)^T E \left(p'_{2,j}\right) = 0$$

给定 5 组对应点，上式可以提供关于本质矩阵 E 的 5 个约束。由于 E 是一个 3*3 的矩阵，而且 E 仅仅在相差一个常数因子下有意义，所以仍需要三个独立约束才能够对 E 进行求解。当 E 确定后， Hartley 等研究表明（Hartley & Zisserman 2000），每个 E 可以分解为 4 种不同的 R, T 组合。所以， 5-点算法旨在从 5 组对应点如何求解本质矩阵 E。这里需要指出的是，5-点算法并不能得到唯一的外参数 R,T，而是有限几组 R,T。

5-点算法的基本思路是：先对上面的线性约束方程组（9 个未知数，5 个约束方程）求解零特征值对应的 4 个特征向量，记为$E_i\ i=1,2,3,4.$则 本质矩阵（展开的向量）E 必为该 4 个向量的线性组合，即：

$$E = xE_1 + yE_2 + zE_3 + wE_4$$

其中x,y,z,w为未知系数。所以，确定本质矩阵 E，就是要确定这些未知系数（如上所述，仅有三个系数独立）。Nister 发现，本质矩阵具有如下二条性质：

（a） 一个本质矩阵具有两个相等的非零奇异值；

（b） 对于一个实数非零的 3×3 的矩阵 E，当且仅当其满足等式 $EE^TE-\frac{1}{2}trace(EE^T)E=0$ 时为本质矩阵。

性质（b）是一个矩阵形式的约束。将线性组合下的 E 代入（b）约束，则可以对系数（x,y,z,w）形成多个非线性约束。Nister 给出了二种求解方法，每种方法均比较复杂，这里不再介绍，感兴趣的学生可阅读原文。5-点算法尽管理论上比较复杂，也不能得到唯一解（最多有 10 组解），但在具体应用中，不正确的解可以通过其它对应点快速剔除，且作者提供了算法代码，所以，目前基于图像的三维重建，人们基本都使用 5-点算法。也许，人们会提的一个问题是，既然 5-点算法得到的多个解可以通过其它对应点剔除，为什么不直接用多组对应点求解本质矩阵呢？这样得到的解的个数会更少。当有 8-组对应点时，理论上可以得到唯一的本质矩阵。这主要是因为实际应用中，二幅图像之间会匹配到大量对应点，但这些对应

点中有很多是错误对应点。为了提高估计的鲁棒性，人们一般在 RANSAC（RANdom SAmplying Consensus ）(Fischler & Bolles 1981) 框架下进行估计，即反复提取待估计问题的最小点集（minimal set），然后对估计的结果进行验证，而 5-组对应点是能够求取本质矩阵的最小点集。

4.3：稀疏三维重建

稀疏三维重建的输入为外极几何图，该图为一个无向图，其顶点为各幅参与重建过程的图像，边连接一对匹配的图像。通常我们认为经几何模型（如本质矩阵、基本矩阵、单应等）过滤后，正确匹配的二维图像特征点仍有足够的个数（如大于 16）。外极几何边包含匹配的图像对之间的二维图像特征点匹配信息以及用于过滤匹配点的几何模型信息。稀疏三维重建的输出为二维图像特征点对应的三维空间点坐标以及在拍摄图像时相机的内外参数。稀疏重建的主要目的是获取场景的三维稀疏点云以及估计相机参数。本节首先介绍在已知相机投影矩阵以及二维特征点匹配时三维空间点坐标的三角化方法，然后对稀疏三维重建中最主要的两类方法：增量式稀疏重建和全局式稀疏重建的原理以及代表性工作进行介绍。

4.3.1：从匹配点恢复空间坐标：三角化方法（triangulation）

在通常情况下，我们采用的相机模型为中心透视投影模型，即由三维空间点射出的光线与二维像平面相交成像。此时，所有光线均穿过同一投影中心 $\boldsymbol{C}$。上述投影过程可通过如下公式进行表述：

$$\boldsymbol{x} \cong \lambda \begin{bmatrix} u & v & 1 \end{bmatrix}^T = \boldsymbol{P}\boldsymbol{X} = \boldsymbol{K}\begin{bmatrix} \boldsymbol{R} & \boldsymbol{T} \end{bmatrix}\boldsymbol{X} = \boldsymbol{K}\begin{bmatrix} \boldsymbol{R} & -\boldsymbol{R}\boldsymbol{C} \end{bmatrix}\boldsymbol{X} \tag{4.1}$$

其中，$\boldsymbol{x}$ 为二维图像投影点的齐次坐标形式，λ 为尺度因子，$\boldsymbol{X}$ 为三维空间点的齐次坐标形式，$\boldsymbol{P}$、$\boldsymbol{K}$、$\boldsymbol{R}$、$\boldsymbol{T}$、$\boldsymbol{C}$ 分别表示相机的投影矩阵、内参矩阵、旋转矩阵、平移向量以及光心位置。

在基于图像的三维重建中，我们的最终目的是通过二维图像上的投影点坐标 $\boldsymbol{x}$ 恢复其对应的三维空间坐标 $\boldsymbol{X}$。然而，对式（4.1）进行简单求逆是不可行的，这是由于任意两个位于视线 $\boldsymbol{y} = \boldsymbol{K}^{-1}\boldsymbol{x}$ 上的点在相差一个尺度因子的意义下是相等的，即：

$$\boldsymbol{y}_1 = \lambda \boldsymbol{y}_2, \lambda \neq 0 \tag{4.2}$$

由于在从三维空间向二维平面的投影过程中深度信息丢失了，因此基于图像的三维重建中的关键挑战之一就是恢复未知的尺度因子 λ。换句话说，我们必须同时恢复沿二维投影点 $\boldsymbol{x}$ 对应的视线 $\boldsymbol{y}$，和从相机光心位置 $\boldsymbol{C}$ 到三维空间点 $\boldsymbol{X}$ 的距离 λ。给出相机的内外参数以及一个二维投影点 $\boldsymbol{x}$ 对应的尺度因子 λ，该点对应的空间点 $\boldsymbol{X}$ 可计算如下：

$$\overline{\boldsymbol{X}} = \lambda \boldsymbol{R}^T \boldsymbol{K}^{-1}\boldsymbol{x} + \boldsymbol{C} \tag{4.3}$$

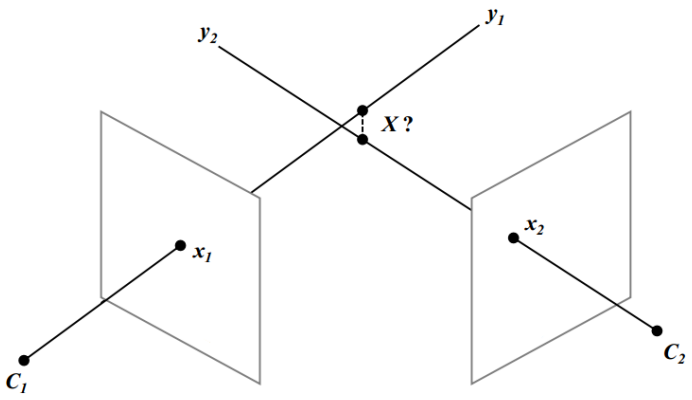

图 4.4：三维空间点的两视图三角化示意图。该过程通过对三维空间点 $\boldsymbol{X}$ 在图像上的二维投影点 $\boldsymbol{x}_1$ 与 $\boldsymbol{x}_2$ 对应的视线 $\boldsymbol{y}_1$ 与 $\boldsymbol{y}_2$ 进行相交实现。

其中，$\overline{\boldsymbol{X}}$ 为 $\boldsymbol{X}$ 的非齐次坐标形式。而当深度 λ 未知时，空间点的位置可通过该空间点在不同相机中的投影点对应的视线相交确定（见图 4.4）。上述视线相交过程称为三角化，该过程可通过直接线性变换（direct linear transformation，DLT）对式（4.1）重新排列进行求解（Hartley& Zisserman, 2000）：

$$\boldsymbol{x} \cong \boldsymbol{P}\boldsymbol{X} = \begin{bmatrix} \boldsymbol{P}_1^T \\ \boldsymbol{P}_2^T \\ \boldsymbol{P}_3^T \end{bmatrix} \boldsymbol{X} \quad \Rightarrow \quad \begin{matrix} \boldsymbol{P}_3^T \boldsymbol{X} u = \boldsymbol{P}_1^T \boldsymbol{X} \\ \boldsymbol{P}_3^T \boldsymbol{X} v = \boldsymbol{P}_2^T \boldsymbol{X} \end{matrix} \quad \Rightarrow \quad \boldsymbol{0} = \begin{bmatrix} \boldsymbol{P}_3^T u - \boldsymbol{P}_1^T \\ \boldsymbol{P}_3^T v - \boldsymbol{P}_2^T \end{bmatrix} \boldsymbol{X} \tag{4.4}$$

上述过程即使在两视图情形下也是超定的，这是因为此时共有 4 个线性等式而未知量共有 3 个。该额外的自由度源于三维空间中两视线通常不会恰好相交。这是由于二维图像投影点 $\boldsymbol{x}$ 上存在测量噪声且投影矩阵 $\boldsymbol{P}$ 估计的不够准确。另需注意的是，当两相机光心重合时上述方程组为奇异方程组。从几何上来讲，此时两视线重合而在此视线上的任意点均为有效解。

通过上述三角化方法能够获取二维图像匹配点所对应的空间点三维坐标的前提是相机已标定好，即相机的投影矩阵 $\boldsymbol{P}$ 是已知的。然而，通常情况下，我们无法预先获知 $\boldsymbol{P}$ 中各元素的值。在这种情况下，我们一般采用从运动恢复结构（structure from motion，SfM）的流程对场景进行稀疏重建，同时获取场景的稀疏表达以及相机的内外参数。稀疏重建主要包括两类方法，增量式稀疏重建和全局式稀疏重建。下文将对该两类方法的原理以及代表性工作进行介绍。

4.3.2：增量式稀疏重建

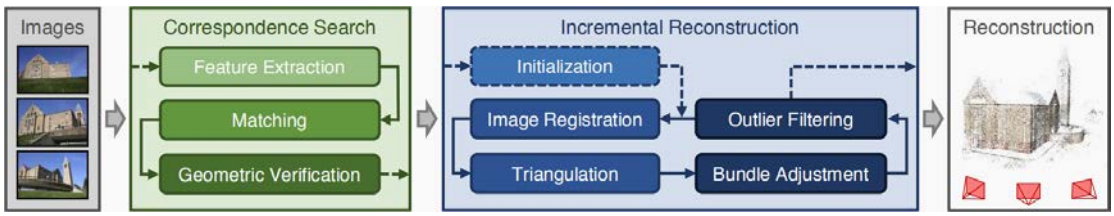

图 4.5：增量式稀疏重建流程图。

增量式稀疏三维重建的主要算法流程如图 4.5 所示，该流程主要包括三个部分：初始重建，增量重建，和全局优化。

初始重建指的是从用于重建的图像集合中选取若干幅图像(通常为 2 幅)作为种子图像，通过估计并分解本质矩阵的方式恢复种子图像之间的相对旋转与相对平移，进而通过三角化获取种子图像稀疏重建的初值，在此基础上通过捆绑调整(bundle adjustment，BA)(Triggs et al., 1999）对种子图像的相机参数以及稀疏点云进行优化，得到初始重建的结果。初始重建在整个增量式稀疏重建的过程中起着重要的作用。 这是由于初始重建结果是后续增量重建的基准，并且如果初始重建结果陷入局部极小值的话，通过后续的优化很难对重建结果进行修正。

增量重建是在种子图像的初始重建的基础之上，依次对其它图像进行重建的过程。该过程一般通过循环迭代的方式进行的。循环内部主要包括三个子步骤：基于透视 n 点（perspective-n-point，PnP）（Gao et al., 2003）的相机定位，基于三角化的场景扩展以及基于捆绑调整的相机参数与场景点云的优化。该迭代过程循环进行直至所有相机均成功定位或者无相机可继续定位。在增量重建的迭代过程中，也存在一个关键问题：以何种顺序添加图像，即下一最优视图（next best view，NBV）（Dunn & Frahm 2009）的选取问题。图像添加顺序也在较大程度上影响着增量式稀疏重建的精度以及鲁棒性。

增量重建过程可得到相机参数与场景稀疏点云的初始估计，将上述结果作为初值，通过全局捆绑调整进一步优化，即可得到增量式稀疏重建的结果。

接下来对增量式稀疏重建的代表性工作进行介绍，主要包括两项工作：Snavely et al. (2008)的方法与 Schönberger & Frahm J. M. (2016)的方法。这两项工作均公开了对应的三维重建开源系统，分别称为 Bundler[1]与 COLMAP[2]，在三维计算机视觉领域有着广泛的应用和重要影响。

Snavely 等(2008)的工作针对从网络上搜索得到的海量图像进行场景重建。下面对这项工作进行具体介绍。

Bundler 算法

首先，在重建初始图像对时，该方法基于如下两个准则进行初始图像对的选取：图像对之间的特征点匹配对个数足够多并且基线（相机光心之间的距离）足够长。满足上述准则的图像对可实现鲁棒的两视图重建。该方法选取初始图像对的具体做法是选取在不能通过单应矩阵对图像特征匹配之间几何关系进行较好描述的图像对集合中，图像特征匹配数量最多的一对图像。单应矩阵可以描述两种几何配置下两幅图像投影点之间的几何关系：场景为一个纯平面，或者两幅图像拍摄时相机光心位置重合（朝向可不同）。因此，如果不能通过一个单应矩阵对两幅图像之间的特征点匹配的几何关系进行有效描述，这说明两相机光心之间有

[1] http://www.cs.cornell.edu/~snavely/bundler/
[2] https://demuc.de/colmap/

一定距离且它们存在公共可见的视场。

具体来说，该方法采用随机抽样一致性（RANdom SAampling Consensus，RANSAC）（Fischler & Bolles 1981）方法估计每对匹配的图像对之间单应矩阵并记录相对于估计的单应矩阵特征点匹配的内点比例。该方法在至少拥有 100 对特征点匹配的图像对中选取内点比例最低的图像对作为初始图像对。

在选取完初始图像对后，该方法通过 5-点算法(Nistér 2004)估计图像对之间的本质矩阵，并对本质矩阵进行分解获取图像对之间的相对旋转与相对平移，然后通过三角化的方式获取初始图像对公共可见点的三维坐标，用作初始三维点集合。在此基础上，该方法采用 SBA 工具包（Lourakis& Argyros 2004）进行捆绑调整，用于优化初始图像对的相机参数以及初始稀疏重建结果。

其次，在添加新图像与空间点时，该方法的图像选取准则是选取能观测到已重建的三维空间点数量最多的相机进行新图像添加。为初始化相机位姿，该方法首先采用基于 RANSAC 的 DLT 算法估计相机投影矩阵 $\boldsymbol{P}$ ，由于 $\boldsymbol{P}$ 有如下形式：

$$\boldsymbol{P} = \boldsymbol{K}[\boldsymbol{R}|\boldsymbol{T}] = [\boldsymbol{KR}|\boldsymbol{KT}] \tag{4.5}$$

因此，矩阵 $\boldsymbol{P}$ 的左边 3×3 子矩阵为一个上三角矩阵 $\boldsymbol{K}$ 与一个旋转矩阵 $\boldsymbol{R}$ 的乘积，可以通过对其进行 RQ 分解来计算相机的内参矩阵 $\boldsymbol{K}$ 与旋转矩阵 $\boldsymbol{R}$。得到 $\boldsymbol{K}$，$\boldsymbol{R}$ 以后，平移向量 $\boldsymbol{T}$ 即为 $\boldsymbol{K}^{-1}\boldsymbol{P}$ 的最右一列。将上述估计的参数值作为初值，通过捆绑调整对相机参数进行优化，在优化过程中，仅允许改变新添加相机的参数，重建模型的其它部分均固定不变。

接下来，该方法将新添加相机观测到的点引入到优化过程。引入新点的条件包括两个：该点至少有两个相机可见；通过对该点三角化可获取该点的良态估计。该方法通过考察用于三角化新添加点所有视线对中夹角的最大值 $\theta_{\max}$ 来评价该点的三角化状况。如果 $\theta_{\max} > 2^{\circ}$，该方法对该点进行三角化并将其引入后续优化过程中。一旦添加了新的点，该方法对目前已重建的整个模型通过捆绑调整进行进一步优化。

图 4.6：Snavely 等(2008) 增量式稀疏重建过程。图中展示的是该方法在 Trevi 数据集上三个不同阶段的结果。左图：初始两视图重建；中图：添加了若干幅图像的中间步骤；右图：采用 360 幅图像得到的最终重建结果(摘自原文)

上述初始化相机、三角化空间点以及捆绑调整优化模型的过程反复迭代进行。每增加一一幅图像，迭代进行一次，直至剩余相机没有足够的可见点为止（该方法中可见点数量需要大于 20）。上述方法的整个过程如图 4.6 所示。需要注意的是，对通过网络搜索得到图像数据集来说，通常能够重建的图像只占参与运算的总图像中的一小部分。

另外，为提升算法鲁棒性以及效率，该方法在上述流程的基础上还做了如下两部分的改进工作：

第一个改进是用来应对二维图像特征点匹配中的不可避免的误匹配现象。误匹配对重建算法有着很大的影响，会导致重建结果出现错误。因此，如何以一种较为鲁棒的方式处理匹配外点十分重要。在每次进行捆绑调整优化过后，该方法通过空间点反投影误差来判断该点对应的图像投影点中是否存在误匹配的情况，并将反投影误差较大的空间点从捆绑调整优化的过程中去除。对于给定图像，用于滤除匹配外点的阈值是与当前图像中的反投影误差的分布相关的。具体来说，该方法首先计算 d_{80}，即当前图像的 80 百分位的反投影误差值，然后将外点阈值设为 $\mathrm{clamp}(2.4d_{80},4,16)$（其中钳止函数 $\mathrm{clamp}(x,a,b)$ 将 x 钳止在区间 $[a,b]$ 之内）。该钳止函数的作用是将在至少一幅可见图像中反投影误差大于 16 像素的空间点记做外点而将在所有可见图像中反投影误差均小于 4 像素的空间点记为内点。在滤除外点后，该方法重新进行捆绑调整并迭代直至检测不到外点为止。

第二个改进是在添加相机时，该方法不再是每添加一个相机就进行优化，而是在添加多个相机之后一起进行优化。在选择要添加的相机时，该方法首先获取与已重建三维空间点对应的二维图像点最多的相机，点数记为 M，然后该方法将所有与已重建三维空间点对应二维图像点数超过 $0.75M$ 的相机均一次进行添加。每次添加多个相机可减少捆绑调整次数，进而提升重建效率。

Schönberger 等的方法

Schönberge & Frahm (2016)方法的基本原理以及主要流程与 Snavely 等 (2008)的方法类似。该工作主要在外极几何图增强，下一最优视图选取以及鲁棒高效三角化三个方面对传统增量式稀疏重建方法进行了改进，下文将对上述三个方面分别进行介绍。

外极几何图增强

该方法提出了一种基于合适几何关系的多模型几何验证策略用来进行外极几何图的增强。首先，该方法进行基本矩阵估计，如果对应该基本矩阵的内点至少有 N_F 个，该方法认为当前图像对通过了几何验证。然后，该方法进行单应估计并获取该单应对应的内点数量 N_H。若 $N_H/N_F<\varepsilon_{HF}$，该方法认为当前图像对对应的两相机之间存在运动(光心不重合)。对于已标定的图像对（内参矩阵 $\boldsymbol{K}$ 已知），该方法还估计了它们之间的本质矩阵并获取该本质矩阵对应的内点数量 N_E。若 $N_E/N_F<\varepsilon_{EF}$，该方法认为给出的标定结果是正确的，在此基础上，该方法分解本质矩阵，三角化匹配内点，进而获取对应点视线夹角中值 α_m。

通过α_m的值可对纯旋转（较小）与纯平面（较大）情形进行区分。根据上述分析，该方法为每个通过几何验证的图像对指定类型标签（常规，纯旋转或纯平面）。由于在进行初始重建时不能利用纯旋转情形的图像对，而更倾向于利用经标定的图像对，上述类型标签有助于初始重建的进行。通过经增强的外极几何图可以较为高效、鲁棒地进行初始重建。另外，该方法不对纯旋转情形的图像对进行三角化以避免退化情况，因此该方法可以提升三角化以及后续图像定位的鲁棒性。

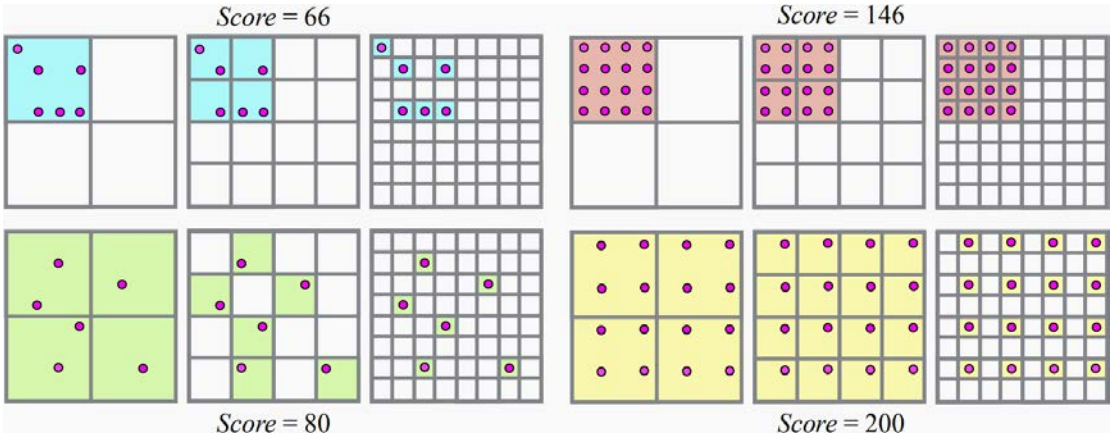

图 4.7：$L=3$时，不同可见点数量（左图与右图）以及不同可见点分布情况（上图和下图）的图像的用于下一最优图像选取的得分（摘自原文）

下一最优视图选取

在进行下一最优视图选取时，该方法认为相对于已重建三维点在二维图像上的可见点数量的准则（Snavely et al., 2008），这些点在图像中的分布形式更为重要。这是由于通常情况下用于相机定位的 2D-3D 点对的数量都是冗余的，而这些点对中二维图像点的较为均匀的分布可使得相机参数估计更为可靠。为此，该方法提出了一种基于多尺度分析的高效下一最优视图选取方法，用于选取可见点数量多且分布均匀的图像。具体来说，该方法先将图像划分为$K_l \times K_l$个图像块。各图像块均有两个状态：空或者满。若其中无可见点，其状态为空；否则为满。为选取包含分布更为均匀的可见点的图像，该方法在进行图像划分时采用的是多尺度金字塔的形式：在金字塔第l（$l=1...L$）层，$K_l=2^l$。用于进行下一最优图像选取的准则为：将每层状态为满的图像块数量进行加权求和，选取得分最高的图像。权重ω_l与所在层数有关：$\omega_l=K_l^2$。图 4.7 给出了不同配置下图像的得分情况。

鲁棒高效三角化

在进行三角化时，若只是按照 4.3.1 节中介绍的方法进行两视图三角化的话，可能会存在由于图像之间基线过短导致三角化结果的不可靠。因此，常见的做法是将图像对之间的特征点匹配通过跨视图连接的方式生成特征点轨迹（feature tracks），并对特征点轨迹进行三角化以获取更为精确的三维点空间坐标。由于特征点轨迹中含有较高比例的匹配外点，该

方法提出了一种高效的、基于 RANSAC 的特征点轨迹三角化方法。在此将特征点轨迹记为 $\mathcal{T}=\{T_n \mid n=1\ldots N_T\}$，其内点率未知。特征点轨迹中的某一观测量包括经归一化的图像点坐标 $\bar{\boldsymbol{x}}_n$（$\bar{\boldsymbol{x}}_n=\boldsymbol{K}^{-1}\boldsymbol{x}_n$）及其对应的相机投影矩阵 $\boldsymbol{P}_n$。该方法的目标是最大化一个良态三角化结果中的观测量支撑集，此处三角化过程通过如下形式表述：

$$\boldsymbol{X}_{ab}\sim\tau\left(\bar{\boldsymbol{x}}_a,\bar{\boldsymbol{x}}_b,\boldsymbol{P}_a,\boldsymbol{P}_b\right)\quad\text{with}\quad a\neq b \tag{4.6}$$

其中，τ 为任选的三角化方法（文中采用 DLT 方法），$\boldsymbol{X}_{ab}$ 为三角化的空间点。所谓良态三角化需满足如下两个条件：首先是足够大的视线夹角 α，α 定义如下：

$$\cos\alpha=\frac{\boldsymbol{T}_a-\boldsymbol{X}_{ab}}{\left\|\boldsymbol{T}_a-\boldsymbol{X}_{ab}\right\|_2}\cdot\frac{\boldsymbol{T}_b-\boldsymbol{X}_{ab}}{\left\|\boldsymbol{T}_b-\boldsymbol{X}_{ab}\right\|_2} \tag{4.7}$$

其次，三角化得到的空间点相对于视图 a 与 b 的深度 d_a 与 d_b 均为正值。另外，若特征点轨迹中的某一观测量反投影误差 e_n 小于给定阈值 t，该方法认为该观测量与三角化结果相符，即属于该三角化结果的观测量支撑集。e_n 通过下式计算：

$$e_n=\left\|\bar{\mathbf{x}}_n-\begin{bmatrix}x'/z'\\y'/z'\end{bmatrix}\right\|_2\quad\text{with}\quad\begin{bmatrix}x'\\y'\\z'\end{bmatrix}=\boldsymbol{P}_n\begin{bmatrix}\boldsymbol{X}_{ab}\\1\end{bmatrix} \tag{4.8}$$

在 RANSAC 过程中进行随机抽样操作时，为提高效率，该方法限定每次抽样产生的最小集合都是唯一的。另外，由于特征点轨迹中的匹配内点比例是未知的，该方法用一个较小的值 ε_0 初始化内点比例，并且当得到更大的一致集时，自适应地对迭代次数 K 进行调整。在此基础上，为进一步提高效率，在迭代过程中，该方法将特征点轨迹中已属于某一致集中的观测量在后续采样时去除。上述迭代过程的终止条件是最新的一致集大小小于 3。

通过上面介绍的增量式稀疏重建的原理及代表性工作可知，这类方法由于在相机增量的过程中通过 RANSAC 进行参数估计，并且通过捆绑调整进行参数优化，使得其对外点的鲁棒性和场景重建精度较高；然而，增量式稀疏重建也存在着较为明显的不足，例如重建结果对初始种子图像选取以及增量过程中对图像添加顺序较为敏感，在进行较大规模场景重建时可能会出现因为累计误差导致的场景漂移现象，并且由于需要反复通过捆绑调整进行参数优化，增量式稀疏重建的计算效率通常较低。

4.3.3：全局式稀疏重建

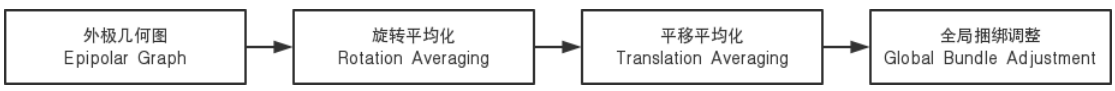

图 4.8：全局式稀疏重建流程图

与增量式稀疏重建对应的还有一类全局式稀疏重建方法。该类方法的主要算法流程如图

4.8 所示。全局式稀疏重建与增量稀疏重建的主要区别是在于初始化相机位姿时，不借助捆绑调整进行优化，而是利用相机之间的相对位姿进行对各个相机的绝对位姿进行估计。在此基础上，通过一次全局捆绑调整对相机位姿以及三维空间点坐标进行优化。全局式稀疏重建主要包括三个部分：旋转平均化（rotation averaging），平移平均化（translation averaging）以及全局捆绑调整优化。

旋转平均化与平移平均化均在外极几何图上进行操作。对于旋转平均化来说，已知相对旋转集合 $\mathcal{R}_{\text{rel}}=\{\boldsymbol{R}_{ij}\}$，对相机的绝对旋转集合 $\mathcal{R}_{\text{global}}=\{\boldsymbol{R}_i\}$ 进行求解。在理想情况下，相对旋转与绝对旋转满足如下关系：

$$\boldsymbol{R}_{ij}=\boldsymbol{R}_j\boldsymbol{R}_i^T \tag{4.9}$$

因此，旋转平均化通常通过求解如下优化问题进行：

$$\arg\min_{\mathcal{R}}\sum_{\boldsymbol{R}_{ij}\in\mathcal{R}_{\text{rel}}}d^{\boldsymbol{R}}\left(\boldsymbol{R}_{ij},\boldsymbol{R}_j\boldsymbol{R}_i^T\right)^p \tag{4.10}$$

其中，$d^{\boldsymbol{R}}(\boldsymbol{R}_1,\boldsymbol{R}_2)$ 为 $SO(3)$ 上的某一距离度量，如角距离、弦距离、四元数距离等（Hartley et al.2013）；$p=1,2$ 表示在距离度量时采用 ℓ_1 或者 ℓ_2 范数作为代价函数。

平移平均化与旋转平均化类似。已知相对平移集合 $\mathcal{T}_{\text{rel}}=\{\boldsymbol{T}_{ij}\}$，对相机的绝对平移集合 $\mathcal{T}_{\text{global}}=\{\boldsymbol{T}_i\}$ 或者相机光心在世界坐标系下的坐标 $\mathcal{C}_{\text{global}}=\{\boldsymbol{C}_i\}$ 进行求解。在理想情况下，相对平移与绝对平移满足如下关系：

$$\lambda_{ij}\boldsymbol{T}_{ij}=\boldsymbol{R}_{ij}\boldsymbol{T}_i-\boldsymbol{T}_j \tag{4.11}$$

因此，对于平移平均化问题，通常通过如下优化问题进行求解：

$$\arg\min_{\mathcal{T}}\sum_{\boldsymbol{T}_{ij}\in\mathcal{T}}d^{\boldsymbol{T}}\left(\boldsymbol{T}_{ij},\frac{\boldsymbol{R}_{ij}\boldsymbol{T}_i-\boldsymbol{T}_j}{\left\|\boldsymbol{R}_{ij}\boldsymbol{T}_i-\boldsymbol{T}_j\right\|}\right)^p \tag{4.12}$$

其中，$d^{\boldsymbol{T}}(\boldsymbol{T}_1,\boldsymbol{T}_2)$ 为 $\mathbb{R}^3$ 上的某一距离度量，如欧氏距离、角距离、弦距离等（Wilson & Snavely 2014）；$p=1,2$ 表示在距离度量时采用 ℓ_1 或者 ℓ_2 范数作为代价函数。需要注意的是，由于平移具有尺度不确定性（λ_{ij}），且通过分解本质矩阵得到的相对平移的精度较差。因此，相较旋转平均化，平移平均化问题更为复杂。

通过旋转平均化与平移平均化的方式对相机的绝对位姿进行估计，可以将外极几何图中的相对位姿误差平均到各个边上，以获取一个从全局来讲较为准确的相机绝对位姿估计。下文中将分别对旋转平均化与平移平均化中的代表性工作进行介绍。

Chatterjee & Govindu (2013)的工作是最具代表性的旋转平均化工作。该方法通过轴角方式表示旋转：$\boldsymbol{\omega}=\theta\boldsymbol{n}$（即轴为向量 $\boldsymbol{n}$，旋转角为 θ 的旋转）。轴角表示与旋转矩阵表示之间的换算关系如下：

$$\boldsymbol{R} = e^{[\boldsymbol{\omega}]_\times} \quad \Leftrightarrow \quad [\boldsymbol{\omega}]_\times = \log(\boldsymbol{R}) \tag{4.13}$$

通过轴角方式表示相机旋转的优势在于，此时相对旋转与绝对旋转之间关系的一阶近似为：

$$\boldsymbol{\omega}_{ij} = \boldsymbol{\omega}_j - \boldsymbol{\omega}_i = \underbrace{\begin{bmatrix} \cdots & -\boldsymbol{I} & \cdots & \boldsymbol{I} & \cdots \end{bmatrix}}_{\boldsymbol{A}_{ij}} \boldsymbol{\omega}_{\text{global}} \tag{4.14}$$

将上述关系对所有已知的相对旋转进行堆叠，可以得到如下表达式：

$$\boldsymbol{A}\boldsymbol{\omega}_{\text{global}} = \boldsymbol{\omega}_{\text{rel}} \tag{4.15}$$

算法 4.1：旋转平均算法

输入：$\mathcal{R}_{\text{rel}} = \left\{ \boldsymbol{R}_{ij}^1, \ldots, \boldsymbol{R}_{ij}^{|\mathcal{R}_{\text{rel}}|} \right\}$

输出：$\mathcal{R}_{\text{global}} = \left\{ \boldsymbol{R}_1, \ldots, \boldsymbol{R}_{|\mathcal{R}_{\text{global}}|} \right\}$

初始化：为 $\mathcal{R}_{\text{global}}$ 赋初值

while $\left\| \Delta\boldsymbol{\omega}_{\text{rel}} \right\| < \varepsilon$ do

1. $\Delta\boldsymbol{R}_{ij} = \boldsymbol{R}_j^{-1}\boldsymbol{R}_{ij}\boldsymbol{R}_i$
2. $\Delta\boldsymbol{\omega}_{ij} = \log(\Delta\boldsymbol{R}_{ij})$
3. 求解 $\boldsymbol{A}\Delta\boldsymbol{\omega}_{global} = \Delta\boldsymbol{\omega}_{rel}$
4. $\forall k \in [1, |\mathcal{R}_{\text{global}}|], \boldsymbol{R}_k = \boldsymbol{R}_k \exp(\Delta\boldsymbol{\omega}_k)$

end while

基于上式，该方法提出了旋转平均化算法，具体流程如算法 4.1 所示。在对算法 4.1 第 3 步的求解过程中，为使算法对外点更为鲁棒，该方法采用的代价函数为 ℓ_1 范数。具体来说，求解算法中第 3 步通过如下优化问题求解：

$$\arg\min_{\Delta\boldsymbol{\omega}_{\text{global}}} \left\| \boldsymbol{A}\Delta\boldsymbol{\omega}_{\text{global}} - \Delta\boldsymbol{\omega}_{\text{rel}} \right\|_{\ell_1} \tag{4.16}$$

算法 4.2：迭代重加权最小二乘算法

初始化：为 $\boldsymbol{x}$ 赋初值

while $\left\| \boldsymbol{x} - \boldsymbol{x}_{\text{prev}} \right\| < \varepsilon$ do

1. $\boldsymbol{x}_{\text{prev}} \leftarrow \boldsymbol{x}$
2. $\Phi \leftarrow \Phi(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})$
3. $\boldsymbol{x} \leftarrow \left(\boldsymbol{A}^T\Phi\boldsymbol{A}\right)^{-1}\boldsymbol{A}\Phi\boldsymbol{b}$

end while

该方法将上述旋转平均化算法命名为 L1RA。在此基础上，该方法为进一步提升旋转平均化的精度，将 L1RA 结果作为初值，通过迭代重加权最小二乘（iteratively reweighted least square，IRLS）方法对旋转平均化的结果进一步优化。通过 IRLS 算法对形如 $\boldsymbol{Ax}=\boldsymbol{b}$ 的问题进行求解的流程如算法 4.2 所示。其中，$\Phi(\boldsymbol{e})$ 为对角阵，其对角元 $\Phi(i,i)=\frac{\sigma^2}{\left(e_i^2+\sigma^2\right)^2}$，$\sigma$ 为调节参数。

算法 4.3：鲁棒旋转平均化算法（L1-IRLS）
L1RA 步骤：
为 $\mathcal{R}_{\text{global}}$ 赋初值
执行算法 4.1，通过式 4.16 求解算法 4.1 中第 3 步
IRLS 步骤：
将 $\mathcal{R}_{\text{global}}$ 值设为 L1RA 步骤的输出
执行算法 4.1，通过算法 4.2 求解算法 4.1 中第 3 步

由于 L1RA 算法对外点鲁棒但精度相对较低，而 IRLS 精度高但结果依赖初值。因此，该方法将两种方法进行结合，将通过 L1RA 方法得到的旋转平均化结果用作初值，采用 IRLS 方法对其进一步优化，该算法流程如算法 4.3 所示，该算法流程命名为 L1-IRLS。

对于平移平均化来说，由于求解更为复杂，相关学者在这方面开展的研究及相应的成果也更多（Moulon et al. 2013; Jiang et al. 2013; Wilson & Snavely 2014; Özyesil & Singer 2015; Cui & Tan 2015 等）。在此，本文对其中的两项具有代表性的工作，Jiang et al. (2013) 与 Cui & Tan(2015)进行介绍。需要指出的是，在进行平移平均化之前，通常已通过旋转平均化方法获取了相机的绝对旋转。

Jiang et al. (2013) 为估计相机光心在世界坐标系下的坐标 $\mathcal{C}_{\text{global}}=\{\boldsymbol{C}_i\}$，该方法首先将两视图内的局部平移 $\mathcal{T}_{\text{rel}}=\{\boldsymbol{T}_{ij}\}$ 变换至全局旋转 $\mathcal{R}_{\text{global}}=\{\boldsymbol{R}_i\}$ 所在的世界坐标系之下：

$$\boldsymbol{C}_{ij}=-\boldsymbol{R}_j^T\boldsymbol{T}_{ij} \tag{4.17}$$

在此基础上，该方法提出了一种通过最小化近似的几何误差的途径对相机光心的绝对位置进行了估计。

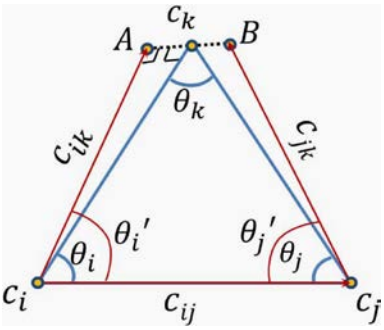

图 4.9：式 4.18 的几何解释（摘自原文）。

首先，该方法对三个相机的特殊情况进行了分析。此时，平移平均化问题转变成了已知三相机光心之间的相对位置关系$\boldsymbol{C}_{ij},\boldsymbol{C}_{ik},\boldsymbol{C}_{jk}$，求解三相机光心的绝对位置$\boldsymbol{C}_i,\boldsymbol{C}_j,\boldsymbol{C}_k$的问题。在理想情况下，三个单位向量$\boldsymbol{C}_{ij},\boldsymbol{C}_{ik},\boldsymbol{C}_{jk}$应当共面，然而，由于存在观测噪声，实际情况下上述三个向量通常并不会共面，即$\left(\boldsymbol{C}_{ij},\boldsymbol{C}_{ik},\boldsymbol{C}_{jk}\right)\neq 0$，在此$\left(\cdot,\cdot,\cdot\right)$表示数量三重积运算。为不失一般性，不妨先假设$\boldsymbol{C}_{ij}$中不存在误差来最小化点$\boldsymbol{C}_k$与直线$l\left(\boldsymbol{C}_i,\boldsymbol{C}_{ik}\right)$以及直线$l\left(\boldsymbol{C}_j,\boldsymbol{C}_{jk}\right)$之间的欧氏距离。在此，$l\left(\boldsymbol{P},\boldsymbol{Q}\right)$表示空间中经过点$\boldsymbol{P}$，朝向为$\boldsymbol{Q}$的一条直线。由于观测噪声的存在，直线$l\left(\boldsymbol{C}_i,\boldsymbol{C}_{ik}\right)$与直线$l\left(\boldsymbol{C}_j,\boldsymbol{C}_{jk}\right)$通常不会共面。因此，在此配置之下$\boldsymbol{C}_k$的最优解为上述两直线的公垂线（图 4.9 中的线段AB）的中点。经推导可知，$\boldsymbol{C}_k$的最优解可通过如下公式近似计算：

$$\boldsymbol{C}_k \approx \frac{1}{2}\left(\left(\boldsymbol{C}_i + s_{ij}^{ik}\left\|\boldsymbol{C}_i-\boldsymbol{C}_j\right\|\boldsymbol{C}_{ik}\right)+\left(\boldsymbol{C}_j + s_{ij}^{jk}\left\|\boldsymbol{C}_i-\boldsymbol{C}_j\right\|\boldsymbol{C}_{jk}\right)\right) \tag{4.18}$$

其中，$\left\|\boldsymbol{C}_i-\boldsymbol{C}_j\right\|$为$\boldsymbol{C}_i$与$\boldsymbol{C}_j$之间的距离；$s_{ij}^{ik}=\sin\left(\theta_j'\right)/\sin\left(\theta_k'\right)=\left\|\boldsymbol{C}_i-\boldsymbol{C}_k\right\|/\left\|\boldsymbol{C}_i-\boldsymbol{C}_j\right\|$与$s_{ij}^{jk}=\sin\left(\theta_i'\right)/\sin\left(\theta_k'\right)=\left\|\boldsymbol{C}_j-\boldsymbol{C}_k\right\|/\left\|\boldsymbol{C}_i-\boldsymbol{C}_j\right\|$分别为基线长度比，角度$\theta_i,\theta_i',\theta_j,\theta_j',\theta_k$的几何意义如图 4.9 所示，角度$\theta_k'$为$\boldsymbol{C}_{ik}$与$\boldsymbol{C}_{jk}$的夹角。式 4.18 关于未知量（$\boldsymbol{C}_i,\boldsymbol{C}_j,\boldsymbol{C}_k$）是非线性的，该方法通过如下方式对其进行线性化。首先，经观察可知：

$$\left\|\boldsymbol{C}_i-\boldsymbol{C}_j\right\|\boldsymbol{C}_{ik}=\left\|\boldsymbol{C}_i-\boldsymbol{C}_j\right\|\boldsymbol{R}_i\left(\theta_i'\right)\boldsymbol{C}_{ij}=\boldsymbol{R}_i\left(\theta_i'\right)\left(\boldsymbol{C}_i-\boldsymbol{C}_j\right) \tag{4.19}$$

其中，$\boldsymbol{R}_i\left(\phi\right)$表示旋转轴为$\boldsymbol{C}_{ij}\times\boldsymbol{C}_{ik}$，旋转角为$\phi$（逆时针）的旋转矩阵。将式 4.19 代入式 4.18 可得：

$$2\boldsymbol{C}_k-\boldsymbol{C}_i-\boldsymbol{C}_j=\boldsymbol{R}_i\left(\theta_i'\right)s_{ij}^{ik}\left(\boldsymbol{C}_j-\boldsymbol{C}_i\right)+\boldsymbol{R}_j\left(-\theta_j'\right)s_{ij}^{jk}\left(\boldsymbol{C}_i-\boldsymbol{C}_j\right) \tag{4.20}$$

其中，$\boldsymbol{R}_j\left(\cdot\right)$表示旋转轴为$\boldsymbol{C}_{ij}\times\boldsymbol{C}_{jk}$的旋转矩阵。类似的，该方法进一步假设$\boldsymbol{C}_{ik}$与$\boldsymbol{C}_{jk}$不存在误差，可得到如下二式：

$$2\boldsymbol{C}_j-\boldsymbol{C}_i-\boldsymbol{C}_k=\boldsymbol{R}_i\left(-\theta_i'\right)s_{ik}^{ij}\left(\boldsymbol{C}_k-\boldsymbol{C}_i\right)+\boldsymbol{R}_k\left(\theta_k'\right)s_{ik}^{jk}\left(\boldsymbol{C}_i-\boldsymbol{C}_k\right) \tag{4.21}$$

$$2\boldsymbol{C}_i-\boldsymbol{C}_j-\boldsymbol{C}_k=\boldsymbol{R}_j\left(\theta'_j\right)s_{jk}^{ij}\left(\boldsymbol{C}_k-\boldsymbol{C}_j\right)+\boldsymbol{R}_k\left(-\theta'_k\right)s_{jk}^{ik}\left(\boldsymbol{C}_j-\boldsymbol{C}_k\right) \tag{4.22}$$

通过联立式（4.20），式（4.21）和式（4.22）进行求解，可以得到三个相机的光心在世界坐标系下的坐标$\boldsymbol{C}_i,\boldsymbol{C}_j,\boldsymbol{C}_k$。

该方法将上述三视图的情况进一步推广到了多视图。具体来说，对于每个三视图均可列出形如式（4.20），式（4.21）和式（4.22）的线性等式。将所有三视图对应的线性等式堆叠到一起，可以得到一个稀疏的齐次线性方程组$\boldsymbol{AC}=\boldsymbol{0}$。其中，$\boldsymbol{C}$为将所有相机光心坐标堆叠到一起的向量，$\boldsymbol{A}$为将所有等式中的未知量的系数堆叠到一起得到的矩阵。上述方程组的解为矩阵$\boldsymbol{A}$的非平凡零向量（non-trival null vector），即矩阵$\boldsymbol{A}^T\boldsymbol{A}$的第四个最小特征值对应的特征向量。该矩阵前三个最小特征值均为0，对应着因世界坐标系原点不确定导致的三个自由度。

上述通过线性方法求解平移平均化问题，在外几何图中存在较多外点时，鲁棒性较差。因此，该方法在进行相机光心位置求解之前，需要先对外极几何图进行了过滤，将误差较大的边进行滤除。具体来说，该方法对各三视图进行平移平均化求解，若通过求解结果反求的相对平移与通过分解本质矩阵得到的原始相对平移偏差大于3°，则认为该三视图不可靠，需进行滤除。另外，该方法还通过旋转平均化结果对外极几何边进行滤除。对于某条外极几何边，若通过旋转平均化结果反求的相对旋转与通过分解本质矩阵得到的原始相对旋转偏差大于5°，则认为该边不可靠，需进行滤除。

Cui & Tan (2015)提出了一种基于相似平均化（similarity averaging）的全局式稀疏重建方法。该方法通过两视图局部重建的方式构建图像的稀疏深度图，在此基础上获取各图像的全局尺度并进一步获取图像对之间的基线长度。在基线长度已知的情况下，平移平均化问题成为了一个适定问题，可以更为简便地进行求解。下文将对该方法的具体流程进行介绍。

该方法主要分为两步：稀疏深度图构建与相似平均化，首先介绍稀疏深度图构建。对于每条外极集合边，该方法根据由本质矩阵分解得到的相对旋转、平移（相对平移向量长度置为 1），采用两视图三角化的方式，实现两视图局部重建。为构建图像i的稀疏深度图，该方法收集图像i参与的所有两视图局部重建结果（最多前 80 个，以两视图特征匹配的数量从大到小排序），并将所有局部重建结果投影至图像i中以构建该图像的稀疏深度图。需要注意的是，由于进行局部稀疏重建时各两视图之间的相对平移向量长度均置为1，因此，为统一深度，需要求解各图像对(i,j)对应的相对尺度s_{ij}^i。具体做法如下：对于图像i中的一个特征点，若其在图像对(i,j)与图像对(i,k)局部重建时均被重建，则这两个局部重建的尺度比可由该特征点的两个重建结果在图像i中的深度比确定：

$$s_{ik}^i\big/s_{ij}^i=d_{ij}\big/d_{ik}:=d_{jk}^i \tag{4.23}$$

对上式两边取对数，可得：

$$\log\left(s_{ik}^{i}\right)-\log\left(s_{ij}^{i}\right)=\log\left(d_{jk}^{i}\right) \tag{4.24}$$

该式为两个局部重建的尺度提供了一个线性等式约束。将关于图像 i 的所有上述等式约束堆叠到一起可得如下非齐次线性方程组：

$$\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b} \tag{4.25}$$

为消除尺度不确定性，该方法将与图像 i 匹配点数最多的图像（此处设为图像 j ）的局部重建尺度（相对平移向量长度）定为 1，即 $\log\left(s_{ij}^{i}\right)=0$ 。该方法采用 ℓ_1 优化的方式对式 4.25 进行求解：

$$\arg\min_{\boldsymbol{x}}\left\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\right\|_{\ell_1} \tag{4.26}$$

通过上述求解过程，可以获取与某幅图像相关的各两视图局部重建的相对尺度，并进而获得各幅图像的稀疏深度图。需要注意的是，此处图像深度图中的深度为相对深度，深度图中点的深度并不是该点的实际深度，但深度图中不同点的深度比值与实际比值相同。

得到各幅图像的稀疏深度图之后，任意一条外极几何边 (i,j) 上的相对关系由满足本质矩阵变为相似变换（而非刚体变换，这是由于得到的深度图为相对深度图导致的）。原则上，该相似变换可通过两视图之间的三维对应关系进行计算，然而，论文作者发现，通过局部重建获得的三维对应点的精度不足以实现高精度的三维对齐。因此，该方法将由本质矩阵分解得到的相对旋转 $\boldsymbol{R}_{ij}$ 与相对平移 $\boldsymbol{T}_{ij}$ 用作上述相似变换中的旋转、平移，而相似变换中的相对尺度由下式计算得到：

$$S_{ij}=s_{ji}^{j}/s_{ij}^{i} \tag{4.27}$$

此时，该方法在各外极几何边 (i,j) 上均求得了一个局部相似变换 $\left(S_{ij},\boldsymbol{R}_{ij},\boldsymbol{T}_{ij}\right)$，进而将通过一个所谓的相似平均化的算法求解各相机的绝对位姿，即已知 $\left\{S_{ij},\boldsymbol{R}_{ij},\boldsymbol{T}_{ij}\right\}$，通过旋转平均化求解各相机的绝对旋转 $\left\{\boldsymbol{R}_i\right\}$，通过尺度平均化求解各相对深度图的尺度因子 $\left\{s_i\right\}$，以及通过基于已知尺度的平移平均化求解各相机的绝对平移 $\left\{\boldsymbol{T}_i\right\}$。其中，该方法采用 Chatterjee & Govindu (2013)方法进行旋转平均化求解。下面将对该方法的尺度平均化以及尺度已知的平移平均化进行介绍。

鲁棒的尺度平均化

为对齐各深度图，该方法为各深度图 D_i 计算一个尺度因子 S_i。根据已知的两两相对尺度，可知：

$$s_i/s_j=S_{ij} \tag{4.28}$$

对上式两边取对数，可得：

$$\log(s_i)-\log(s_j)=\log(S_{ij}) \tag{4.29}$$

对于每条外极几何边，均可获得一个类似上式的等式，将所有等式放到一起，可以得到如下非齐次线性方程组：

$$\boldsymbol{A}_s\boldsymbol{x}_s=\boldsymbol{b}_s \tag{4.30}$$

在求解上式时，为消除尺度不确定性，该方法将第一幅图像的尺度因子置为 1，即 $\log(s_1)=0$。然后，该方法通过 ℓ_1 优化的方式对式 4.30 进行求解：

$$\arg\min_{\boldsymbol{x}_s}\left\|\boldsymbol{A}_s\boldsymbol{x}_s-\boldsymbol{b}_s\right\|_{\ell_1} \tag{4.31}$$

尺度已知的平移平均化

在各深度图的尺度因子已知的前提下，两视图 (i,j) 之间的基线长度可通过如下方式计算：

$$\lambda_{ij}=\frac{1}{2}\left(s_i s_{ij}^i+s_j s_{ij}^j\right) \tag{4.32}$$

在此基础上，可得到如下关于相机光心绝对位置的线性等式：

$$\boldsymbol{R}_j\left(\boldsymbol{C}_i-\boldsymbol{C}_j\right)=\lambda_{ij}\boldsymbol{T}_{ij} \tag{4.33}$$

对于每条外极几何边，均可获得一个类似上式的等式，将所有等式放到一起，可以得到如下非齐次线性方程组：

$$\boldsymbol{A}_c\boldsymbol{x}_c=\boldsymbol{b}_c \tag{4.34}$$

为消除世界坐标系原点位置的不确定性，该方法将第一个相机光心置于世界坐标系原点，即 $\boldsymbol{C}_1=\boldsymbol{0}$。然后，该方法通过 ℓ_1 优化的方式对式 4.34 进行求解：

$$\arg\min_{\boldsymbol{x}_c}\left\|\boldsymbol{A}_c\boldsymbol{x}_c-\boldsymbol{b}_c\right\|_{\ell_1} \tag{4.35}$$

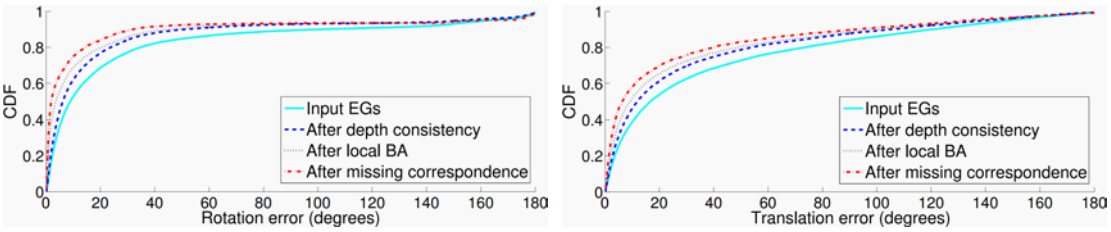

图 4.10：Gendarmenmarkt 数据集上相对旋转与相对平移的累计误差分布曲线（cumulative distribution function，CDF）。此数据集的原始外极几何图中相对旋转与相对平移均包含较大误差。经该方法中的深度一致性检验，局部捆绑调整以及缺失对应点分析后，外极几何图中的相对旋转与相对平移的精度均得到了提升（摘自原文）。

该方法通过上述流程实现了平移平均化。另外，由于全局式平移平均化算法对外极几何图中的外点敏感，因此，该方法还对外极几何图进行了如下操作以提高其精度：深度一致性检验，局部捆绑调整以及缺失对应点分析。各项操作均可对外极几何图的精度有一定程度上

的提升，具体如图 4.10 所示。

全局式稀疏重建通过相机之间的相对位姿恢复各相机的绝对位姿，将误差平均分配到外极几何边上，因此不存在增量式稀疏重建的累计误差问题；另外，全局式稀疏重建仅需一次全局捆绑调整，因此其重建效率更高。然而，全局式稀疏重建，尤其是在进行平移平均化时，对外点较为敏感；另外，由于在旋转平均化以及平移平均化过程中可能会存在误过滤外极几何边的操作，从而会产生丢失图像的情况。

4.4：稠密三维重建

稠密三维重建的输入为稀疏三维重建的输出，即为二维图像特征点对应的三维空间点坐标以及在拍摄图像时相机的内外参数。稠密三维重建的输出为通过图像间的稠密匹配（逐像素匹配）获得的场景的三维稠密点云。稠密重建一直是计算机视觉领域的一个研究热点，近年来涌现了很多方法。为实现稠密重建，通常采用的是多视图立体视觉（multi-view stereo，MVS）技术，其基本原理是在已知相机位姿的前提下，通过图像一致性（photo-consistency）函数，例如平方误差和（sum of squared differences，SSD），绝对误差和（sum of absolute differences，SAD）以及归一化互相关（normalized cross correlation，NCC）等，实现图像间的稠密匹配，进而重建场景的稠密三维点云。根据 Seitz 等 (Seitz et al. 2006) 的分类原则，稠密三维重建算法可大致分为四类：基于体素的方法，基于表面进化的方法，基于特征点扩散的方法以及基于深度图融合的方法。上述四种方法中，基于特征点扩散的方法与基于深度图融合的方法由于更适用于大规模场景的稠密三维重建，因此应用更为广泛。其中，基于特征点扩散的方法的基本原理是基于初始稀疏点云，采用迭代的方式，通过最小化图像一致性函数优化新点的参数（位置、法向等），实现点云的扩散；基于深度图融合的方法的基本原理是首先计算每幅图像对应的深度图，然后将各深度图进行融合得到稠密点云。下文将对这两种方法的代表性工作进行介绍。

4.4.1：基于特征点扩散的稠密重建

Furukawa & Ponce (2010)是基于特征点扩散的稠密重建方法中最具有代表性的工作。该方法作者将代码开源并命名为 patch-based multi-view stereo（PMVS）。PMVS 在三维计算机视觉领域有着广泛的知名度与应用。本文首先介绍该方法用到的一些基本要素，在此基础上，对该方法的算法流程进行介绍。

该方法用到的二个基本要素为：平面块（patch）模型与图像模型。

平面块模型

平面块 p 本质上是局部切平面的近似，其几何性质由中心 $\boldsymbol{c}(p)$ 与法向 $\boldsymbol{n}(p)$ 唯一确定。因此，与通常的只将点的空间位置作为图像一致性函数的输入不同，该方法设计的图像一致性函数的输入包括点的位置与法向，因此更为鲁棒。基于定义的图像一致性函数，重建平面

块的过程即为通过优化平面块参数（位置、法向）来最大化该函数的过程。初看之下，该函数有 5 个待优化参数，这是由于空间位置与法向的自由度分别为 3 和 2。然而，在进行平面块优化的过程中，由于平面块不应沿表面切向运动而只应沿表面法向运动，因此，平面块优化过程中仅有 3 个待优化参数（平面块法向以及沿法向的位置偏移量）。

图像模型

通过平面块进行表面表示具有较高的灵活性，然而，由于这种表示方式缺少连接性信息，因此很难查找或获取相邻平面块。该方法通过借助重建的平面块在其可见图像中的投影实现平面块邻居的查找与访问。具体来说，该方法对每幅图像 I_i 进行网格划分，划分的网格记为 $C_i(x,y)$，其大小为 $\beta\times\beta$ 像素（该方法中 $\beta=2$），如图 4.11 所示。给出一个平面块 p 及其可见图像集合 $V(p)$，该方法将 p 投影至 $V(p)$ 中的各幅图像上。然后，该方法获取投影至各图像网格 $C_i(x,y)$ 的平面块集合 $Q_i(x,y)$。在此基础上，平面块的邻接性可通过平面块在其可见图像上的投影所在网格的邻接性进行判断。

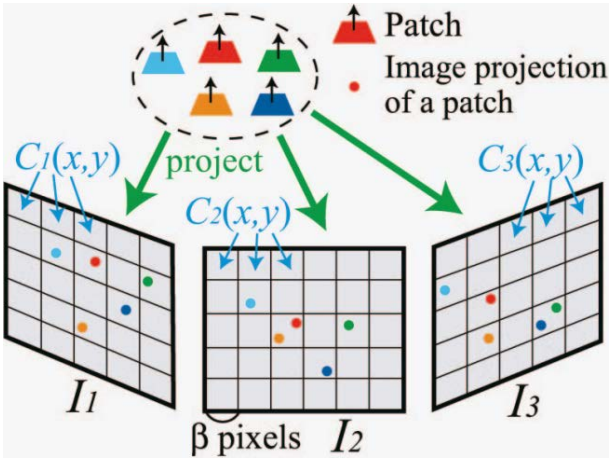

图 4.11：重建的平面块在其可见图像中的投影以及可见图像网格划分示意图（摘自原文）

接下来对该方法的算法流程进行介绍。该方法主要包含三步：初始特征匹配、平面块扩展和平面块过滤。其中，初始特征匹配步骤的目的是生成稀疏的平面块集合；而平面块扩展与过滤步骤是迭代进行的（通常迭代 3 次），用以生成更为稠密的平面块集合以及过滤错误的平面块。上述三步将在下文中具体介绍。

算法 4.4：初始特征匹配算法算法

输入：各图像中检测的特征点

输出：初始稀疏平面块集合 $\mathcal{P}$

初始化：$\mathcal{P}\leftarrow\varnothing$

For 各幅以 $O(I)$ 为光心的图像 I

For 各个图像 I 中检测得到的特征点 f

$\mathcal{F} \leftarrow \{$满足极线一致性的特征点$\}$

将 $\mathcal{F}$ 中的特征点按照其（三角化后）与 $O(I)$ 距离从小到大排序

For $\mathcal{F}$ 中的各特征点 f'

通过式（4.36），（4.37）初始化 $\boldsymbol{c}(p)$，$\boldsymbol{n}(p)$

通过图像视角差异初始化 $V(p)$

优化 $\boldsymbol{c}(p)$，$\boldsymbol{n}(p)$

根据与 I 的图像一致性更新 $V(p)$

If $|V(p)| < \gamma_v$

回到最里面的 For 循环

将 p 投影至各可见图像并加入对应图像网格的平面块集合 $Q_j(x,y)$

将 p 加入到重建平面块集合 $\mathcal{P}$

退出最里面的 For 循环

初始特征匹配

该方法第一步采用高斯差分（Difference-of-Gaussian，DoG）以及 Harris 算子（Szeliski 2010）在各输入图像中检测局部特征。对于图像 I_i 中检测到的每个特征点 f，该方法在其它图像上收集距 f 在这些图像上对应的外极线小于 2 像素的同类型（DoG 或 Harris）特征点集合 $f' \in \mathcal{F}$，并对各特征点对 (f, f') 进行三角化得到对应的三维点。然后，该方法将这些三维点作为平面块可能的中心并按照其与图像 I_i 的光心 $O(I_i)$ 之间的距离从小到大排序。在此基础上，该方法通过如下方式实现初始特征匹配。给出一对特征点匹配 (f, f')，该方法首先通过如下方式初始化其对应的潜在平面块：

$$\boldsymbol{c}(p) \leftarrow \{\text{对}(f, f')\text{进行三角化}\} \tag{4.36}$$

$$\boldsymbol{n}(p) \leftarrow \overrightarrow{\boldsymbol{c}(p)O(I_i)} \Big/ \left|\overrightarrow{\boldsymbol{c}(p)O(I_i)}\right| \tag{4.37}$$

在进行上述操作时需要预先知道当前平面块的可见图像集合 $V(p)$（图像数量一般为 5），其通过图像 I_i 邻近图像与 I_i 的视角差异确定。给定平面块参数初值，即可通过最大化图像一致性函数对平面块参数进行优化。优化过后，$V(p)$ 中仅含有对于平面块 p 来说，与图像 I_i 的图像一致性数值大于给定阈值的可见图像。如果 $|V(p)| \geq \gamma_v$，通常 $\gamma_v = 3$，该方法将 p 保留。上述步骤的算法流程如算法 4.4 所示。

算法 4.5：平面块扩展算法

输入：初始稀疏平面块集合 $\mathcal{P}$

输出：扩展后的平面块集合 $\mathcal{P}$

While $\mathcal{P} \neq \varnothing$

从 $\mathcal{P}$ 中选取并删掉一个平面块 p

For 每个含有 p 的图像网格 $C_i(x,y)$

通过式 4.38 确定一个用于扩展的网格集合 $\mathbf{Cells}(p)$

For $\mathbf{Cells}(p)$ 中的每个网格 $C_i(x',y')$

初始化平面块 p'：$\boldsymbol{n}(p') \leftarrow \boldsymbol{n}(p)$，$V(p') \leftarrow V(p)$，$\boldsymbol{c}(p')$ 射线与平面相交

优化 $\boldsymbol{c}(p')$，$\boldsymbol{n}(p')$

更新 $V(p')$

If $|V(p')| < \gamma_v$

回到最里面的 For 循环

将 p' 加入 $\mathcal{P}$ 中

将 p' 加入对应的 $Q_i(x,y)$ 中

平面块扩展

该步骤的目标是在每个图像网格 $C_i(x,y)$ 中至少重建一个平面块。为实现上述目标，该方法迭代地从已有的平面块出发，在其附近的空白空间生成新的平面块。具体来说，给出一个平面块 p，该方法首先通过下式获取目前仍不含任何平面块的图像邻居网格 $\mathbf{Cells}(p)$：

$$\mathbf{Cells}(p) = \left\{ C_i(x',y') \middle| p \in Q_i(x,y), Q_i(x',y') = \varnothing, |x-x'| + |y-y'| = 3 \right\} \tag{4.38}$$

对于 $\mathbf{Cells}(p)$ 中的每个图像网格 $C_i(x,y)$，该方法通过如下扩展流程生成新的平面块 p'。首先，该方法通过 $\boldsymbol{n}(p)$，$V(p)$ 初始化 $\boldsymbol{n}(p')$，$V(p')$。对于 $\boldsymbol{c}(p')$ 来说，该方法将其初始化为平面块 p 所在的平面与过 $C_i(x,y)$ 中心的视线的交点。然后，该方法对 $\boldsymbol{c}(p')$，$\boldsymbol{n}(p')$ 进行优化并基于优化后的 $\boldsymbol{c}(p')$，$\boldsymbol{n}(p')$ 对 $V(p')$ 进行更新。最终，若 $|V(p')| \geq \gamma_v$，该方法认为此时平面块 p' 重建成功并更新其可见图像的 $Q_i(x,y)$。上述过程迭代进行，直至各平面块重建完成。上述步骤的算法流程如算法 4.5 所示。

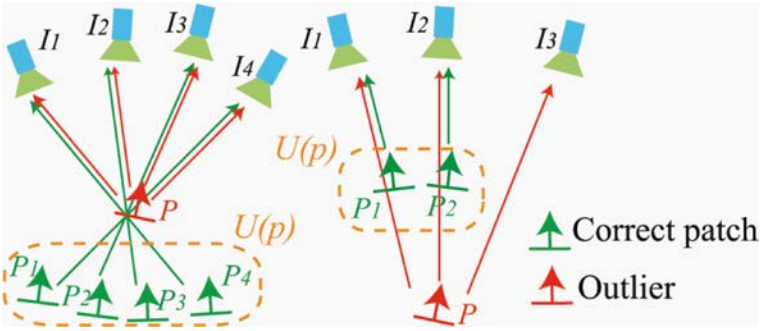

图 4.12：通过全局可见一致性来滤除外点（图中红色平面块）。图中（左图与右图）由 p_i 指向 I_j 的箭头表示 $I_j \in V(p_i)$，即平面块 p_i 在图像 I_j 中可见；$U(p)$ 表示与 p 可见性不一的平面块集合（图摘自原文）

平面块过滤

上文中的平面块扩展步骤仅依赖于图像一致性度量来重建平面块。因此，该步骤难以避免生成错误的平面块。针对这种情况，该方法最后采用如下两个过滤步骤以滤除错误的平面块。第一个过滤方法依赖于可见一致性。为实现该过滤方法，该方法首先定义平面块 p 与 p' 的邻接性：满足如下约束的平面块对 p 与 p' 为邻居平面块：

$$\left|(\boldsymbol{c}(p)-\boldsymbol{c}(p'))\cdot\boldsymbol{n}(p)\right|+\left|(\boldsymbol{c}(p)-\boldsymbol{c}(p'))\cdot\boldsymbol{n}(p')\right|<\gamma_d \tag{4.39}$$

其中，γ_d 为判断阈值。该方法将 $p'\in U(p)$ 记为与平面块 p 可见性不一致的平面块集合，即平面块对 p 与 p' 不是邻居平面块，但却在 p 的某些可见图像中存于同一个图像网格内（见图 4.12）。这样的话，若如下不等式成立，该方法认为 p 为外点并将其滤除：

$$|V(p)|(1-\mathcal{C}(p))<\sum_{p_i\in U(p)}1-\mathcal{C}(p_i) \tag{4.40}$$

其中，$\mathcal{C}(p)$ 为 p 的图像一致性得分均值。直观上来讲，若 p 为外点，$|V(p)|$ 与 $1-\mathcal{C}(p)$ 都应该较小，此时将 p 滤除。

对于第二种过滤方法，该方法引入了一种正则化操作：对于每个平面块 p，首先获取在其所有可见图像 $V(p)$ 中 p 所在或相邻的图像网格中所有的平面块。如果获取的平面块中 p 的邻居（根据式 4.39 获得）所占比例低于 0.25，该方法认为 p 为外点并将其滤除。

基于特征点扩散的方法可以生成精度较高、分布较为均匀的点云。然而，这类方法在弱纹理区域容易造成扩散空洞；而且，由于这类方法的流程串联性较强，难以进行并行化操作，因此相对来说计算效率较低。

4.4.2：基于深度图融合的稠密重建

Shen (2013)为一项比较有代表性的基于深度图融合的稠密重建方法，该方法主要分为三步：立体对选取，深度图计算与深度图融合，其流程图如图 3.30 所示。下文将对这三步

逐一进行介绍。

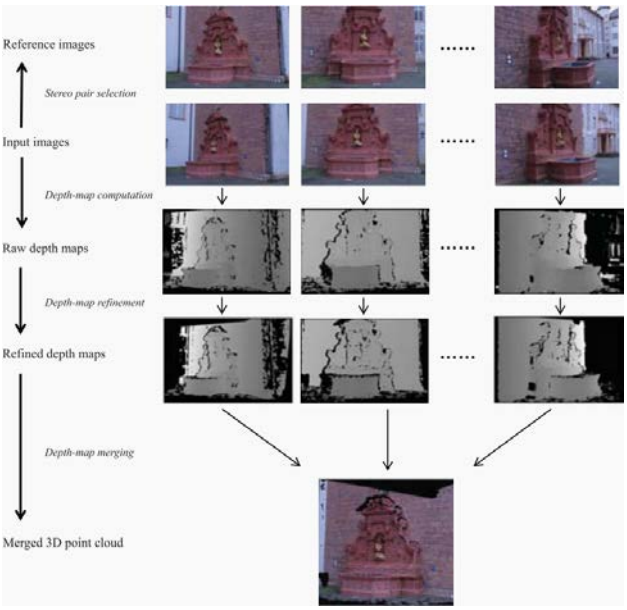

图 4.13：Shen (2013)算法流程图

立体对选取

对于图像集合中的每幅图像，该方法选取一幅参考图像用于立体计算。立体对的选取对于立体匹配以及最终的稠密重建的精度都有比较重要的影响。

该方法采用类似 Li 等(Li et al. 2010)的方法选取合适的立体对。假设当前共有 n 幅图像，对于第 i 幅图像，该方法计算 $\theta_{ij}, j=1,\ldots,n$，即图像 i 与图像 j 朝向的夹角以及 $d_{ij}, j=1,\ldots,n$，即图像 i 与图像 j 相机光心的距离。对于满足 $5°<\theta_{ij}<60°$ 的图像，该方法计算对应的 d_{ij} 的中值 $\bar{d}$ 并将 $d_{ij}>2\bar{d}$ 或者 $d_{ij}<0.05\bar{d}$ 的图像滤除。该方法将经上述滤除过程仍保留下来的图像做为图像 i 的邻居图像，记做 $N(i)$。$N(i)$ 中 $\theta_{ij}\cdot d_{ij}$ 值最小的图像作为图像 i 的参考图像。

深度图计算

对于每对选取的立体对，该方法采用与 Bleyer 等 (Bleyer et al.2011)类似的思想计算深度图。该方法的关键思想是在目标图像的各像素上找到好的支撑平面，此时聚合匹配代价（aggregated matching cost）最小（见图 4.14 左图）。

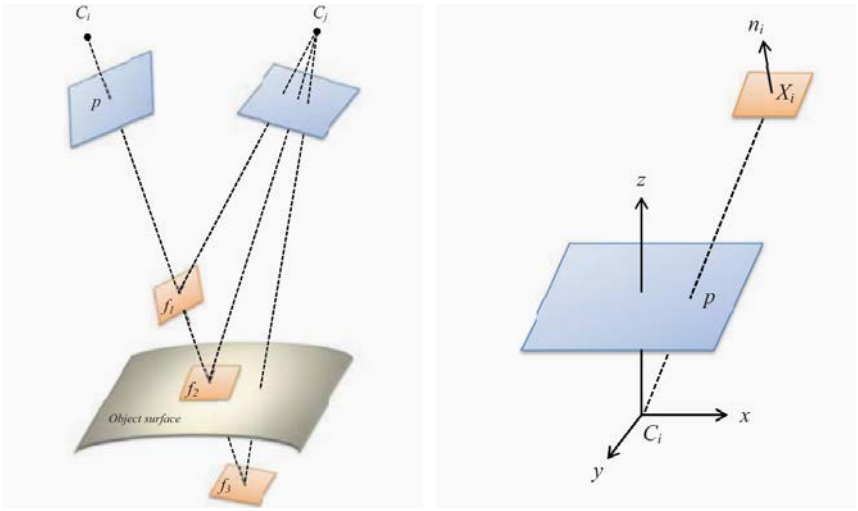

图 4.14：左图，对于目标图像中的每个像素 p，该方法估计其对应的三维平面。C_i 与 C_j 分别为目标图像与参考图像的相机光心，f_1，f_2 与 f_3 为 p 对应视线上的三个三维平面。显然其中 f_2 有着最小的匹配代价。右图，由目标相机 C_i 坐标系下的三维点 $\boldsymbol{X}_i$ 及其法向 $\boldsymbol{n}_i$ 表示的支撑平面。其中，C_i 为目标图像的相机光心，$C_i - xyz$ 为该相机的相机坐标系（摘自原文）。

该方法将通过目标相机坐标系下的一个三维点 $\boldsymbol{X}_i$ 及其法向 $\boldsymbol{n}_i$ 表示三维空间中的平面，如图 4.14 右图。

给出图像对 I_i 与 I_j 及其对应的相机参数 $\{\boldsymbol{K}_i, \boldsymbol{R}_i, \boldsymbol{C}_i\}$ 与 $\{\boldsymbol{K}_j, \boldsymbol{R}_j, \boldsymbol{C}_j\}$，该方法首先为目标图像 I_i 中的各像素 p 分配一个随机的三维平面。假设 p 的齐次坐标为：

$$[u \quad v \quad 1]^T \tag{4.41}$$

由于表示该随机平面的三维点 $\boldsymbol{X}_i$ 位于 p 对应的视线上，该方法从深度有效范围 $[\lambda_{\min}, \lambda_{\max}]$ 上随机选取一个深度 λ，那么 $\boldsymbol{X}_i$ 可通过下式计算：

$$\boldsymbol{X}_i = \lambda \boldsymbol{K}_i^{-1} p \tag{4.42}$$

然后，该方法采用相机 C_i 的球坐标为该平面随机指定法向：

$$\boldsymbol{n}_i = \begin{bmatrix} \cos\theta \sin\phi \\ \sin\theta \sin\phi \\ \cos\phi \end{bmatrix} \tag{4.43}$$

其中，θ 为一个在 $[0°, 60°]$ 之间的随机角度，ϕ 为一个在 $[0°, 360°]$ 之间的随机角度。上述角度范围的设置是基于如下假设：若一个平面块在图像 I_i 中可见，则其法向 $\boldsymbol{n}_i$ 与该相机坐标系 z 轴的夹角应低于某一阈值（该方法中该阈值设为 $60°$）。

为根据平面 $f_p = \{\boldsymbol{X}_i, \boldsymbol{n}_i\}$ 计算像素 p 的聚合代价，该方法首先计算由该平面诱导的单应

矩阵：

$$\boldsymbol{H}_{ij}=\boldsymbol{K}_j\left(\boldsymbol{R}_j\boldsymbol{R}_i^T+\frac{\boldsymbol{R}_j\left(\boldsymbol{C}_i-\boldsymbol{C}_j\right)\boldsymbol{n}_i^T}{\boldsymbol{n}_i^T\boldsymbol{X}_i}\right)\boldsymbol{K}_i^{-1} \tag{4.44}$$

然后，该方法在像素 p 周围设定一个正方形窗口 W 。对于 W 中的每个像素 q ，该方法通过式 4.44 得到的单应矩阵 $\boldsymbol{H}_{ij}$ 计算其在参考图像中对应的像素位置。基于此，像素 p 的聚合匹配代价 $m\left(p,f_p\right)$ 定义为：1-q 与 $\boldsymbol{H}_{ij}\left(q\right)$ 之间的归一化互相关得分：

$$m\left(p,f_p\right)=1-\frac{\sum_{q\in W}\left(q-\bar{q}\right)\left(\boldsymbol{H}_{ij}(q)-\overline{\boldsymbol{H}_{ij}(q)}\right)}{\sqrt{\sum_{q\in W}\left(q-\bar{q}\right)^2\sum_{q\in W}\left(\boldsymbol{H}_{ij}(q)-\overline{\boldsymbol{H}_{ij}(q)}\right)^2}} \tag{4.45}$$

经上述初始化以后，目标图像 $\boldsymbol{I}_i$ 中的每个像素均对应着一个随机的三维空间平面。然后，该方法对图像 $\boldsymbol{I}_i$ 中的各个像素逐一优化其对应的三维平面，共进行三次迭代。第一次迭代中，该方法从左上角像素开始，按行主序遍历，直至到达右下角像素。第二次迭代中，该方法翻转方向，由右下角像素开始，按行主序遍历，到左上角像素结束。第三次迭代中遍历顺序与第一次一致。

在每次迭代中，该方法对各像素进行如下两种操作：空间传播与随机分配。其中，空间传播操作用于比较和传播当前像素与邻居像素的空间平面。在第一和第三次迭代中，邻居像素为当前像素的左边以及上边的相邻像素，而在第二次迭代中，邻居像素为当前像素的右边及下边的相邻像素。将当前像素 p 的邻居像素记为 p_N ，其对应的三维平面记为 f_{p_N} 。若 $m\left(p,f_{p_N}\right)<m\left(p,f_p\right)$ ，该方法将 f_{p_N} 传播至当前平面，即 $f_p\leftarrow f_{p_N}$ 。上述传播过程的依据机理为：相邻像素很可能有着相似的对应三维平面。

对于每个像素 p ，当空间传播操作完成后，该方法采用随机分配操作对平面 f_p 进一步优化。具体步骤为：给出参数变化范围 $\{\Delta\lambda,\Delta\theta,\Delta\phi\}$，1)选取一个随机的平面参数 $\{\lambda',\theta',\phi'\}$ 其中， $\lambda'\in[\lambda-\Delta\lambda,\lambda+\Delta\lambda],\theta'\in[\theta-\Delta\theta,\theta+\Delta\theta],\phi'\in[\phi-\Delta\phi,\phi+\Delta\phi]$；2）通过式 4.42 与 4.43 计算新的平面参数 $f_p'=\{\boldsymbol{X}_i',\boldsymbol{n}_i'\}$；3）若 $m\left(p,f_p'\right)<m\left(p,f_p\right)$， $f_p\leftarrow f_p'$；4）将 $\{\Delta\lambda,\Delta\theta,\Delta\phi\}$ 减半；5）回到第 1)步。上述优化过程迭代 10 次，该方法中 $\Delta\lambda=\frac{\lambda_{\max}-\lambda_{\min}}{4},\Delta\theta=15°,\Delta\phi=90°$ 。

深度图融合

在深度图计算完成后，该方法通过对深度图进行融合以获取完整的三维模型，该方法中的融合过程与 Tola 等方法 (Tola et al.2012)类似。对于图像 $\boldsymbol{I}_i$ 中的每个像素 p ，若其匹

配代价$m(p, f_p)$小于 0.2，该方法通过该点的深度以及相机参数将该点反投影至三维空间：

$$\boldsymbol{X} = \lambda \boldsymbol{R}_i^T \boldsymbol{K}_i^{-1} \boldsymbol{p} + \boldsymbol{C}_i \quad (4.46)$$

然后，该方法将点 $\boldsymbol{X}$ 投影至 I_i 的邻居图像 $N(i)$。假设 I_j 为 $N(i)$ 中的第 j 个邻居图像，该方法将点 $\boldsymbol{X}$ 相对于相机 j 的深度定义为 $d(\boldsymbol{X}, j)$，将 $\boldsymbol{X}$ 在图像 I_j 投影处根据图像 I_j 的深度图得到的深度定义为 $\lambda(\boldsymbol{X}, j)$。若 $\boldsymbol{X}$ 在 I_j 中的匹配代价小于 0.2 且 $\frac{|d(\boldsymbol{X}, j) - \lambda(\boldsymbol{X}, j)|}{\lambda(\boldsymbol{X}, j)} < 0.01$，该方法认为 $\boldsymbol{X}$ 与 I_i 以及 I_j 是一致的。当 $N(i)$ 中至少有 2 幅图像与 $\boldsymbol{X}$ 已知，该点保留。最终，所有保留的三维点即为最终用于表示场景的三维点云。

基于深度图融合的方法由于可以并行计算深度图，因此这类方法更适用于大规模场景的稠密重建；另外，由于这类方法是通过融合深度图生成稠密点云，因此得到的点云数量较大。然而，这类方法的重见效果在很大程度上依赖于邻居图像组的选择。

4.5：整体优化：捆绑调整 （bundle adjustment）

捆绑调整是三维重建中广泛使用的一类非线性优化方法。目前文献中已有大量开源算法软件，这里不再进行介绍。感兴趣的同学可参阅相关文献，如 Triggs (1999)。

参考文献

Bleyer M. et al. (2011). Patchmatch stereo—stereo matching with slanted support windows. In British Machine Vision Conference, pp. 14.1-14.11.

Chatterjee A. & Govindu V. M. (2013). Efficient and robust large-scale rotation averaging. In IEEE International Conference on Computer Vision, pp. 521-528.

Cui Z. & Tan P. (2015). Global structure-from-motion by similarity averaging. In IEEE International Conference on Computer Vision, pp. 864-872.

Dunn, E., & Frahm, J. M. (2009). Next best view planning for active model improvement. In British Machine Vision Conference, pp. 1-11.

Fischler M. A. & Bolles R. C. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM, vol. 24, no. 6, pp. 381-395.

Furukawa Y. & Ponce J. (2010). Accurate, dense, and robust multiview stereopsis. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 8, pp. 1362-1376.

Gao X. et al. (2003). Complete solution classification for the perspective-three-Point problem. IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 25, no. 8, pp.930-943.

Hartley R. I. & Zisserman A. (2000). Multiple view geometry in computer vision. Cambridge University Press.

Hartley R. I. et al. (2013). Rotation averaging. International Journal of Computer Vision, vol.

103, no. 3, pp. 267-305.

Jiang N. et al. (2013). A global linear method for camera pose registration. In IEEE International Conference on Computer Vision, pp. 481-488.

Li J. et al. (2010). Bundled depth-map merging for multi-view stereo. In IEEE Conference on Computer Vision and Pattern Recognition, pp. 2769–2776.

Lourakis M. & Argyros A. (2004). The design and implementation of a generic sparse bundle adjustment software package based on the Levenberg-Marquardt algorithm. Technical Report 340, Institute of Computer Science-FORTH, Heraklion, Greece.

Moulon P. et al. (2013). Global fusion of relative motions for robust, accurate and scalable structure from motion. In IEEE International Conference on Computer Vision, pp. 3248-3255.

Nistér D. (2004). An efficient solution to the five-point relative pose problem. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 6, pp. 756-770.

Özyesil O. & Singer A. (2015). Robust camera location estimation by convex programming. In IEEE Conference on Computer Vision and Pattern Recognition, pp. 2674-2683.

Pollefeys et al. (1998). Self-calibration and metric reconstruction in spite of varying and unknown international camera parameters, ICCV199

Triggs B. et al. (1999). Bundle adjustment — a modern synthesis. In International Workshop on Vision Algorithms, Springer-Verlag, pp. 298-372.

Schönberger J. L & Frahm J. M. (2016). Structure-from-motion revisited. In IEEE Conference on Computer Vision, pp. 4104-4113.

Seitz S. M. et al. (2006). A comparison and evaluation of multi-view stereo reconstruction algorithms. In IEEE Conference on Computer Vision and Pattern Recognition, pp. 519–528.

Shen S. (2013). Accurate multiple view 3D reconstruction using patch-based stereo for large-scale scenes. IEEE Transactions on Image Processing, vol. 22, no. 5, pp. 1901-1914.

Shen S. & Hu Z. (2014). How to select good neighboring images in depth-map merging based 3D modeling. IEEE Transactions on Image Processing, vol. 23, no. 1, pp. 308-318.

Snavely N. et al. (2008). Modeling the world from internet photo collections. International Journal of Computer Vision, vol. 80, no. 2, pp. 189-210.

Szeliski R. (2010). Computer vision: Algorithms and applications. Springer-Verlag New York, Inc.

Tola E. et al. (2012). Efficient large-scale multi-view stereo for ultra high-resolution image sets. Machine Vision Applications, vol. 23, no. 5, pp. 903-920.

Wilson K. & Snavely N. (2014). Robust global translations with 1DSfM. In European Conference on Computer Vision, pp. 61-75.