

指导教师： 杨涛

提交时间： 2016/3/14

CVPR2015 Paper Translation

No: 01

姓名： 吕咸强

学号： 2013300219

班号： HC001310



三维重建的直接结构估计

摘要

大对数传统运动恢复结构技术 (SFM) 的任何结构场景计算需要摄像机的姿态估计。在这项工作中我们发现, 当结合单/多单应性估计, 一般的欧式刚性约束提供了一个简单的场景恢复方法, 不需要明确的摄像机位姿计算。这种直接的结构估计开启了一个新的设计 SFM 系统的方法, 这种方法使得结构和运动估计的顺序发生了反转。我们发现, 这种替代的方法在横向运动平面或一般的人造场景情况下, 可以很好地用于重建的场景结构和摄像机位姿。

第一部分——简介

结构运动 (SFM) 是计算机视觉中的一个经典问题, 已积极研究了几十年。近年来, 由于不断增长的工业应用需求, 如导航, 机器人技术和电影/游戏制作, 在推进系统的可靠性和可扩展性方面的应用技术方面已取得重大进展。几乎所有现代的 SFM 系统都是基于两个或三视图之间特征对应的相对位姿估计。这些相关的位姿在之后将被合并成一个整体的坐标系统。然后连同所有经过束调整的相机参数, 计算细化场景结构。因此, 可靠和准确的相对位姿估计是一个强大的 SFM 系统的关键。然而, 计算相对可靠的位姿是一个不平凡的任务。大多数技术受制于平面场景引起的不稳定性, 这是常见的人造环境中经常发生的。因此, 在 SFM 系统经常采用一个单独的步骤来检测显性单应性。另一方面, 平面结构本身实际上提供了一个强大的几何约束, 可以用来提高重建的质量。

先进的 SFM 系统除了这个著名局限外, 李 [14] 指出一般的 SFM 方法存在一个技术的盲点。几乎所有的传统培养方法, 摄像机运动估计永远是第一位的, 其次是三维结构的计算,

李 [14] 所从事的研究最早提出了绕过运动估计的实际方法。

当我们在传统的 SFM 体系下评价一些基本原理, 如理论的优雅和实用性, 我们对在实际的 SFM 系统结构优先方法的可行性和优势感兴趣。事实上, 我们发现, 在已知的摄像机内部参数的前提下, 在不同两幅图像里面的同一个三维点的深度比可以从单应性有关的两个图像点直接推断。此外, 欧氏刚度约束意味着 2 个三维点之间的欧氏距离是不变的刚体变换。当通过不同视图结合上述单应诱导的深度比, 我们可以从刚性约束, 推导出一个简单的方程来求解两个三维点的相对深度, 只有获得至少三种不同的视图, 即场景结构可以直接确定了一个共同的规模从三标定图像。值得注意的是, 虽然我们的方法包括单应估计, 我们不需要点分享的对应性, 我们也不从单应矩阵估计任何相机参数。为方便参考, 我们称这种方法为直接结构估计 (DSE), 并将以下使用它。

评价 DSE 在实际 SFM 系统的应用潜力, 我们从估计的结构求得了相机的参数。具体而言, 我们计算的场景结构中的每个图像的摄像机视图, 用一个简单的三维刚体变换方法获得相机的相对位姿。使用非线性优化如束调整, 可以进一步精确估计结果。我们发现, 该方法尤其适合侧向运动, 不受限于现有的平面结构的数量。在实际应用中, 这确实是一个理想的特性, 由于横向运动有利于结构计算, 并且在数据采集的三维重建时很普遍。

第二部分——相关工作

我们的工作可以被看作是一个“结构第一” SFM 技术的例子。对比的丰富 SFM 技术文献资料, 这是一个相对“空缺”的部分。早期的作品已经推测无需显式计算摄像机运动获得一般的三

维结构方法的可行性。基于图的刚性理论，李第一次提出了实际实施这样的课题的方法，在嵌入的三维点的实际坐标前，计算一个子集的点间的欧氏距离。然而，将这样的方案一个强大的和可扩展的SFM系统不明显。点间的距离已被用于在早期的视觉工程，以获得多视图不变量。我们的方法也利用刚体变换下的点间距离的不变性得到的约束。Tardif等人利用分解框架，提出了一种可以恢复场景结构的基础约束，主要用于仿射相机。阿利亚加等人。[1]提出了一种结构估计方案通过从SFM制定消除运动参数，但它需要进行非线性平差问题的初始化。

众所周知，先验知识的场景平面可以大大方便三维重建问题。基于摄像机自标定和无标定视图的三维重建已在文献[28]，[11]，[3]研究。周等人提出了一个完全自动化的SFM系统基于非校准的视频序列场景的主平面检测。虽然这些作品处理未标定图像，从多 π 的观点旨在同时恢复摄像机的运动和场景结构，我们证明从图像对应单应性和作为一个通用的可扩展的SFM系统组件的关系中直接估计场景结构的可行性。

我们的DSE方法涉及鲁棒多单应检测，这是一个具有挑战性的和活跃的研究课题。经典问题的目标是集群的图像点，使得它们形成一个最小数量的平面区域和每个区域准确地覆盖尽可能多点。然而，我们的目标是略有不同的经典问题声明。事实上，我们并不关心是否单应检测数量是最优的，并且不关心可以同时有多个对应的作业点。这种放松使我们的问题更容易，我们提出了一个简单的方法来实现我们的目标。

我们使用三个视图为DSE的基本构建块。相对的姿态从场景结构计算很容易地集成到现有的应用系统，如[10]，[18]。特别是，我们使用

江等提供的开源代码，登记在全球范围内的摄像机，并申请获得最终的重建。

第三部分——直接结构估计

下文中，我们首先介绍了校准相机单应诱导结构约束。这个约束使我们能够得知从2个不同的视图观察的三维点的深度比率。然后，我们将使用这个深度比得出的方程，用于解决在同一视图中观察到的两个三维点的相对深度。在一般情况下，我们得出的方程可以求出2个有效的解。因此，我们建议使用第三个视图来解决这个歧义，并求得一对三维点的唯一解。我们还设计了一个鲁方案来重现场景结构，通过所有这样的成对点估计，这些成对点通常会被噪声和错误污染。

3.1 单应诱导结构约束

我们先对单应诱导结构进行形式化证明。如果一对相应的校准点 $p = (x, y, 1)^T$ 和 $p' = (x', y', 1)^T$ 在图片 I 和 I' 通过单应矩阵 H 相关。我们能得到下面的等式：

$$\lambda p' = Hp, \quad (1)$$

式中 λ 是一个标量

假设有 $H = R + \frac{tn^T}{d_\pi}$ ，在摄像机坐标系中， t 表示相机旋转和平移的两个视图之间的旋转和平移， X 表示在摄像机坐标系中的视图平面。在这里， n 表示平面的法线， d_π 公式表示平面 X 到相机中心视图 I 的距离。

命题：

假设 d 和 d' 表示一个三维点 X 在视图 I 和 I' 中的深度，通过预计的二维点 p 和 p' ，相对的，我

们得到公式 $\lambda = \frac{d'}{d}$ 。

证明：

假设 X 标记 π 平面的三维点，并且

满足平面等式 $\mathbf{X}^T \mathbf{n} - d_\pi = 0$ 。相机关于视图 I 和 I' 的投影矩阵分别由 $[I \ 0]$ 和 $[R \ t]$ 提供。给出了 X 点在视图 $I(I')$ 中的深度 $d(d')$ ，

我们有等式：

$$d\mathbf{p} = \mathbf{X}, \quad (2)$$

$$d'\mathbf{p}' = \mathbf{R}\mathbf{X} + \mathbf{t}. \quad (3)$$

用平面等式替换等式 2 和等式 3 有：

$$\frac{1}{d} = \frac{\mathbf{n}^T \mathbf{p}}{d_\pi}, \quad (4)$$

$$\frac{d'}{d} \mathbf{p}' = \mathbf{R}\mathbf{p} + \frac{\mathbf{t}}{d}. \quad (5)$$

联合等式 4 等式 5，我们得到等式：

$$\frac{d'}{d} \mathbf{p}' = \left(\mathbf{R} + \frac{\mathbf{t}\mathbf{n}^T}{d_\pi}\right)\mathbf{p} = \mathbf{H}\mathbf{p}, \Rightarrow \lambda = \frac{d'}{d}.$$

3.2 恢复相对深度

在下文中，我们通过单应诱导深度比和欧氏刚性约束推导，得到求解三维点对的相对深度的简单公式。

给定两对相关点 $(\mathbf{p}_i, \mathbf{p}'_i)$ 和 $(\mathbf{p}_j, \mathbf{p}'_j)$ ，我们根据欧式

刚性约束，用 (d_i, d'_i) 和 (d_j, d'_j) 表示在视图 I 和 I' 的相对深度。在任何刚体变换下，三维点 \mathbf{X}_i 和 \mathbf{X}_j 之间的距离并没有改变，有：

$$\|d'_i \mathbf{p}'_i - d'_j \mathbf{p}'_j\| = \|d_i \mathbf{p}_i - d_j \mathbf{p}_j\|$$

从单应性有关的每一对相关点的得到深度比率 $\lambda_i = \frac{d'_i}{d_i}$ 和 $\lambda_j = \frac{d'_j}{d_j}$ ，通过简单的数学变换，我们能得到下面的等式：

$$\|\lambda_i \frac{d'_i}{d_j} \mathbf{p}'_i - \lambda_j \mathbf{p}'_j\| = \|\frac{d_i}{d_j} \mathbf{p}_i - \mathbf{p}_j\|.$$

让 $\alpha = \frac{d_i}{d_j}$ 我们能得到下面关于 α 的二次方程式：

$$A\alpha^2 + B\alpha + C = 0, \text{ where} \quad (8)$$

$$A = \|\lambda_i \mathbf{p}'_i\|^2 - \|\mathbf{p}_i\|^2,$$

$$B = -2(\lambda_i \lambda_j \mathbf{p}'_i{}^T \mathbf{p}'_j - \mathbf{p}_i{}^T \mathbf{p}_j),$$

$$C = \|\lambda_j \mathbf{p}'_j\|^2 - \|\mathbf{p}_j\|^2.$$

我们能够很容易的求解方程式，得到 α 的两个不同解。相当于我们得到了三维点 X_i 和 X_j 在视图 I 的相对深度。

给定第三个视图 I'' ，可以得到另外两个 α 的解，我们可以选择一个解，来满足两个等式并且是正解。事实上我们直接解决以下最小化问题，以获得最佳的解决方案：

$$\alpha_{ij} = \arg \min_{\alpha} |A_1 \alpha^2 + B_1 \alpha + C_1| + |A_2 \alpha^2 + B_2 \alpha + C_2|. \quad (9)$$

我们使用 α_{ij} 去表示 \mathbf{X}_i 和 \mathbf{X}_j 的相对深度的估计值。这里 (A_1, B_1, C_1) 和 (A_2, B_2, C_2) 分别是视图对 (I, I') 和 (I, I'') 计算的系数。

事实上，等式 9 分别最小化了 X_i 和 X_j 通过相机视图 I, I' 和 I'' 测量的欧式距离的不同。

3.3 结构估计

到目前为止，我们已经展示了如何获得两三维的相对深度点，通过给出在三视图中的对应性和关联的单应性。理想的情况下，一个可以固定的任意点的深度，并计算其余的场景结构得到，最终得到一个整体的结构。实际情况下，结果显然会因选定的参考点产生偏差。由于 α 的计算简单、易于并行化对于不同的 3D 点对，我们详尽地计算所有的三维点，这些三维点在三视图里面是相互对应的。

现在我们表示视图 I 中的和另外两个视图有相关性的点集为 $S = \{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_N\}$ 并且用 N

表示这样的点的总数。对于每一个属于点集的点 ($\mathbf{p}_i \in S$) 并且点的深度固定为 $d_i = 1$ 的点, 在同一个视图里面的

的所有点的深度可以表示为 $\alpha_{ki} = \frac{d_k}{d_i}$ 。

如果有零噪音在数据中, 我们有:

$$\{\alpha_{ki}\} = \beta_{ji}\{\alpha_{kj}\}, \forall k \in [1, N],$$

意味着每一个点的深度仅仅通过全局缩放因子来区别。在掺杂噪音的情况下, 每一个点对

$(\alpha_{ki}, \alpha_{kj})$ 会估算得到一个不同的 β_{ji} 。因此我们用 RANSAC (预期 β_{ji} 的阈值被设定为

1%) 来计算深度的平均的缩放因子。

在应用每一个点估计的缩放因子后,

只需对每一点 \mathbf{p}_i 的深度估计进行相同的计算。

在实践中, 我们会计算视图 I' 的 α'_{ij} , 把这个值继承到视图 I , 通过视图之间点的深度比应用单应性公式: $\tilde{\alpha}_{ij} = \alpha'_{ij} \frac{\lambda_j}{\lambda_i}$, 同样的视图 I'' 也会

重复这个计算。我们计算所有 $\tilde{\alpha}_{ij}$ 的平均值作为每一对点 α_{ij} 的估计值。如果 $\tilde{\alpha}_{ij}$ 的标准偏差较大, 为了更好地鲁棒性, 我们会丢弃这些 $\tilde{\alpha}_{ij}$ 值。图片 3(b) and (c) 展示了一个我们恢复的典型结构。没有估计任何相机的移动参数, 我们很好地重现了场景中的主要平面结构。具有正交关系的两墙立面之间 castle-p30 完好的保存下来。

3.4 多应性估计

在应用 DSE 之前, 我们需要检测单应性的存在, 并且从图像的对应关系计算各单应变换。这是一个典型的多模型拟合的问题, 针对这项任务提出了先进的算法, 例如[27], [9]。然而, 我们的要求是略有不同的, 相比于经典问题相对宽松。首先, 我们不在乎

单应发现的最佳数量只要每个点群真正符合单应性。其次, 每个图像点可以归属到不同的单应性。这是我们所期望的事实, 因为在两视图每个点单应矩阵拟合会给出一个估计的深度比。我们还想指出一般局部平面单应处处存在, 单应性是许多几何非平面结构的很好的近似, 如果局部的匹配。因此, 我们采用一个简单的“匹配-优化”的方法扫描间隔均匀的图像区域, 进行单应性检测。最近的工作[23]也使用了类似的策略, 以产生立体匹配的平面假设。

具体来说, 我们首先将参考视图的图像平面进行分割, 分成非重叠的 $L \times L$ 框。然后我们在每一个图像框中使用 RANSAC 方法去产生单应性假设, 使用所有的相关点在第二副图像里面进行的内/离群检测(我们设定阈值为 2 个像素点对于一副 1600×1200 的图像)。我们认定一个单应性假设为正确只有正确数超过 10 个。对每一个正确的假设, 我们重复进行所有的正确单应性拟合, 直到没有发现更多新的内点, 进行内/离群点测试。我们优化的单应性保证桑普森的错误最小。通过‘优化’的这一步, 非局部内点也可以聚集。这对空间无关的几何平面的表面或建筑外立面是有用的。我们展示了在图 3(一)所提出的技术的例子结果。

在我们的例子中, 一个图像点可以参与一个以上的。通常情况下, 在我们测试的例子中的, 对每个图像点的单应矩阵拟合范围从 0 到 6。个人的深度比估计和平均值之间的差异通常小于 1%。多点的单应性估计更精确和稳定, 我们的体重每一个公式用于单应估计集数, 以加权平均计算。我们通过用于单应性估计的内点的数量权衡每一个 λ 值的权重, 然后求权重的平均值用于我们计算。

第四部分——SFM 系统的集成

DSE 最直接的应用是作为一个一般的 SFM 系统构建块。给一个图像的集合，我们可以对每三个有充分重叠的图像应用 DSE 法处理（通过考虑他们之间共同对应相关性的数目）。通过计算一个三维刚性转换之间的场景结构恢复，计算每三个图像间的相对姿态。这些相对的姿态很容易满足 SFM 系统的要求如 [18]，[10]。我们将在下面详细描述每一步。

一旦我们从单应检测和 DSE 获得场景结构，我们可以恢复相机姿态，作为三维刚体变换使用 SVD [2] [5] 的副产品。考虑到 2D-3D 的对应关系，我们没有使用标准的相机绝对姿态算法（例如 EPNP [12]）因为 PnP 算法也存在在一个单一的平面结构存在不稳定性。事实上，我们发现的三维刚体变换给出了类似的结果，在一般的相机姿态估计。在实践中，为了达到最佳效果，我们可以使用这些初始相机姿态确定剩余的图像的对应，不是在 DSE 步骤恢复的，还能完善了相机的姿势通过。

给定的超过三幅图像，我们为每个三重视图计算场景结构，首先恢复相机的相对位姿。然后，我们将这些相对位姿与江等提出的算法进行匹配，完成多幅图像的三维重建。

第五部分——实验

我们使用合成数据和真实数据来评价 DSE，为了充分了解其表现和在 SFM 应用的潜力。对合成的平面结构的场景 (S)，我们比较了 DSE 和有代表性的三个校准相对位姿算法。就是说，我们选择直接单应矩阵分解算法 (HD) [16]，5 点算法 (2v5p) 对核面几何 [19] 和四点算法 (3V 4P) [20] 对三焦张量。为简单起见，我们只测试 HD 算法，在一个单一的平面结构情况下。

对于合成实验，我们遵循传统的建立如前面的文献中所描述的，例如

[19]，[20]。如图 4 所示，第一台相机是面向世界坐标系的。二次相机放置在距离第一相机 0.1 个单位处，第三相机安放在前两个相机中间的基线之间。相机的平移方向是由角 θ 控制，对应的横向运动 $\theta = 90^\circ$ 。第二和第三相机的光学轴旋转时，它的光轴穿过成像点的质心，并且保证 x 轴平行于 x-z 平面，y 轴指向相同的半空间作为世界的 Y 轴。摄像机的水平视场为 45° ，图像分辨率为 352×288 。我们首先对图像坐标的零均值高斯噪声有不同的标准偏差。

现场都在用最少的 0.5 个单位 1 和场景深度单位深度摄像头视景生成。在平面场景的情况下，飞机是这样的，它通过现场的视锥和正常的偏离角度 0 轴 30 度中心产生。在多个平面的情况下，平面方向产生类似于单平面的情况下。每架飞机的位置是由它指定场景中任意一点确定截。随机采样的场景点被任意投射到这些平面的可见部分。我们产生五个不同的飞机的测试。注意它是不容易模拟现实的一块明智的平面场景，而不引入偏置。我们只有在 DSE 算法的行为感兴趣，完美的点聚类给出单应安装在所有情况下。在我们的计算所涉及的所有参数，仍然估计从给定的噪声数据。

5.1 准确性和稳定性

我们测试了在不同的的噪声水平下和不同摄像机平移方向，DSE 的精度和稳定性。所有的平移方向中，正向和侧向运动是两个特殊的运动，经常在数据采集中遇到的。然而，作为经验法则，用于重建的目的，向前运动通常是不建议的场景点的三角剖分，可以非常敏感的图像噪声和小摄像机运动误差。然而，在我们的实验中也观察到了这种缺点，在我们的实验中，给出另一种解释。因此，对横向运动的算法的性能是重建的目的更多的在实践中的重要性。我们使用公开可用

的源代码 2v5p, 和自己比较的 3V 4P 和 HD 的实现。由于 HD 算法放弃两有效分解 [16], 我们对结果通过寻找共同的平面法线的参考视图和其他的两个视图之间的两个单应性恢复。

摄像机位姿估计精度的比较在图 5 和图 6 中给出。这些数据中的每个数据点计算了超过 100 个试验。在前两个列中分别显示了相对旋转误差和平移误差。这些错误是计算在内的姿态估计的相对误差小于 3° 旋转。在第三列中给出的比例的相机姿态估计与总误差。通过这样做, 我们有一个更好地了解如何良好的算法是在获得正确的解决方案, 其实际的数值稳定性, 图像噪声。

我们发现, 2v5p 和 3v4p 算法在图像噪声的存在下的平面场景很不稳定 (图 5 (c)), 尤其是在快速运动的情况 (图 5 (a))。另一方面, 无论可用性的数量, DSE 方法不断产生的横向运动最好的结果, (图 5 (a) 和图 6 (a))。HD 算法表现出良好的性能, 当侧向运动, 图像噪声低时, 但当图像噪声增加, 选择正确的平面变得更为困难。然而, DSE 在向前运动, 性能但不符合其侧向运动性能。当图像噪声变大时, 该算法难以在正确的结构上进行求解。我们观察到相同的行为与高清。一个更仔细的检查, 我们发现像点的位移通常可以通过正面平行平面时, 相机经历了向前的运动具有相对小的基线诱导单应解释清楚。所有 DSE 产生的异常运动, 因为这个特别“假结构”。我们给出这样的例子在图 7 (1)。

当场景包含多个层面, 2v5p 产生各种运动效果良好, 优于 3V 4P。DSE 对侧向运动最好的结果, 但它仍然受到了运动结构混乱。原因类似于单平面的情况。

目前 DSE 独立计算每个点 0 最佳相对深度不考虑与其他点的一致性。我们相信这一全球一致性是解决向前

运动的结构混乱的关键。因此, 一个有趣的未来方向是考虑的所有点之间的相对深度的一致性和选择的配置, 最大限度地减少了所有的图像观测的重新投影误差。然而, 我们可以看到从图 5 (c) 和图 6 (c), 舒适的操作区的电流范围从 50° 到 90° 无论场景结构的类型。我们认为这是一个互补的算法性能比标准的相对位姿的算法如 2v5p。

5.2 三维重建

当将 DSE 整合到一个 SFM 系统如 [10], 我们需要消除假三重重建。在这里, 三重验证显然行不通因为虚假结构和相机姿态通常共同构成了模糊的解决方案。相反, 我们进行两两验证。每个视图对, 我们可以比较它们之间的相对位姿估计从不同的三视图来识别异常。我们只是考虑一个相对位姿 (因此三重态的来源) 作为一个离群值, 如果其最小旋转差从至少 2 个其他的解决方案是大于 3° 。我们的测试基于六真实图像序列的基于 DSE 的 SFM 系统。

我们使用的基准数据集提供一个定量的评价。使用 SIFT [15] 计算对应特征值。三视图的计算, 三视图是首先连接每个图像与其他图像对应生成三大部分, 然后收集所有的这些视图对形成三视图。结果见表 1。我们测试我们的算法与地面真教标 (GT) 和校准读取 EXIF 标签 (EXIF), 在这里, $R3err$ 和 $t3err$ 分别表示平均相对转动误差和平移误差度在三视图内。在旋转的绝对平均误差 (度) 和相机的位置 (ERN) 在最后的 BA 处理前, 通过 $Rerr$ 和 $Cerr$ 分别给出。绝对的相机位置误差在最终 BA 后由给出 $Cerr(BA)$ 。DSE 得到重建 (SVD) 表 1 使用 EXIF 校准后在图 8 中可以看到。为参考的目的, 我们也列出了使用 [10] 算法获得的结果。

有趣的是, 我们产生类似的结果, 以 [10] 对这些基准数据集, 即使没

有应用三视图。特别是，我们得到一个更好的相机构成的“castle-p30”的初始化。该数据集包含图像的平面建筑立面为主，5点算法产生的图像对之间的相对位姿估计误差较大。总的来说，三图像的BA应适用于初始相机姿态获得DSE(SVD)确保最后的BA最佳的初始化。

我们比较两个三维重建得到的DSE基于SFM(三视图BA)和原始点方法[10]直观图9。“街”的序列具有38的图像，我们收集105个按时间来避免重复结构混乱造成连接相邻的图像。“建筑”的例子有67个图像，我们收集193个。我们使用三视图于两个方法。我们可以清楚地看到从图9到5点算法由于噪声的相对位姿估计，产生错位的重建[10]。基于DSE的SFM给出了更好的结果，虽然它没有进行相对位姿估计在几个视图‘建筑’的行为相比传统的相对位姿算法打开思考一个强大的SFM系统设计的新方法。我们相信有基于DSE方案改进的SFM充足的房间。例如，可以考虑解决相对深度在全局范围内提高其性

鸣谢：这项研究是由新加坡ADSC肝癌的研究机构的研究经费资助，技术研究（一星级）

参考文献：

1. D. G. Aliaga, J. Zhang and M. Boutin
"Simplifying the reconstruction of 3d models using parameter elimination"
Proc. ICCV, pp. 1-8,
2. K. S. Arun, T. S. Huang and S. D. Blostein
"Least-squares fitting of two 3-d point sets"
IEEE Trans. PAMI, no. 5, pp. 698-700, 1987
3. A. Bartoli and P. Sturm
"Constrained structure and motion from multiple uncalibrated views of a piecewise planar scene"
IJCV, vol. 52, no. 1, pp. 45-64, 2003

例子。这很可能是由于附近的小基线向前运动的存在。我们还展示了一个重建的“店方”序列包含122幅图像，通过我们基于DSE的SFM图10。

我们目前不合理的matlab实现的DSE大约需要8秒，在一个典型的图像组大小1600×1200分段平面场景为主的2.53hz CPU。由于所有的计算涉及DSE是轻量级的，增速是微不足道的。

第六部分——总结、将来工作

在这项工作中，我们表明，给定的三个校准的图像和场景和可检测的平面，我们可以直接估计结构，而不计算任何摄像机运动参数。这一有趣的发现导致了SFM的方案，是建立在相反的顺序，即，首先计算结构，随后的姿态估计。实验结果表明，该结构的计算特别适合快速运动的场景结构的类型。这种互补的算法能，并利用单应检测和拟合线处理室内环境。最后，结合DSE和传统的相对位姿估计达到最大的稳定性和通用性，是一个有趣的研究课题。

4. D. Crandall, A. Owens, N. Snavely and D. Huttenlocher
"Discrete-continuous optimization for large-scale structure from motion"
Proc. CVPR, pp. 3001-3008,
5. D. W. Eggert, A. Lorusso and R. B. Fisher
"Estimating 3-d rigid body transformations: a comparison of four major algorithms"
Machine Vision and Applications, vol. 9, no. 5, pp. 272-290, 1997
6. M. A. Fischler and R. C. Bolles
"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, 1981
7. N. M. Grzywacz and E. C. Hildreth
"Incremental rigidity scheme for recovering structure from motion: Position-based versus

- velocity-based formulations"
J. Opt. Soc. Am. A, vol. 4, no. 3, pp. 503-518, 1987
8. R. Hartley
"In defense of the eight-point algorithm"
IEEE Trans. PAMI, vol. 19, no. 6, pp. 580-593, 1997
9. H. Isack and Y. Boykov
"Energy-based geometric multi-model fitting"
IJCV, vol. 97, no. 2, pp. 123-147, 2012
10. N. Jiang, Z. Cui and P. Tan
"A global linear method for camera pose registration"
Proc. ICCV, pp. 481-488,
11. R. Kaucic, R. Hartley and N. Dano
"Plane-based projective reconstruction"
Proc. ICCV, vol. 1, pp. 420-427,
12. V. Lepetit, F. Moreno-Noguer and P. Fua
"Epnnp: An accurate $o(n)$ solution to the pnp problem"
IJCV, vol. 81, no. 2, pp. 155-166, 2009
13. M. Lhuillier and L. Quan
"A quasi-dense approach to surface reconstruction from uncalibrated images"
IEEE Trans. PAMI, vol. 27, no. 3, pp. 418-433, 2005
14. H. Li
"Multi-view structure computation without explicitly estimating motion"
Proc. CVPR, pp. 2777-2784,
15. D. G. Lowe
"Distinctive image features from scale-invariant keypoints"
IJCV, vol. 60, no. 2, pp. 91-110, 2004
16. Y. Ma, S. Soatto, J. Kosecka and S. S. Sastry
"An invitation to 3-d vision: From images to geometric models (interdisciplinary applied mathematics)" 2005
17. D. Martinec and T. Pajdla
"Robust rotation and translation estimation in multi view reconstruction"
Proc. CVPR, pp. 1-8,
18. P. Moulon, P. Monasse and R. Marlet
"Global fusion of relative motions for robust, accurate and scalable structure from motion"
Proc. ICCV, pp. 3248-3255,
19. D. Nister
"An efficient solution to the five-point relative pose problem"
IEEE Trans. PAMI, vol. 26, no. 6, pp. 756-770, 2004
20. D. Nister and F. Schaffalitzky
"Four points in two or three calibrated views: Theory and practice"
IJCV, vol. 67, no. 2, pp. 211-231, 2006
21. M. Pollefeys, D. Nister, J.-M. Frahm, A. Akbarzadeh, P. Mordohai, B. Clipp, C. Engels, D. Gallup, S.-J. Kim and P. Merrell
"Detailed real-time urban 3d reconstruction from video"
IJCV, vol. 78, pp. 143-167, 2008
22. L. Quan
"Invariants of six points and projective reconstruction from three uncalibrated images"
IEEE Trans. PAMI, vol. 17, no. 1, pp. 34-46, 1995
23. S. N. Sinha, D. Scharstein and R. Szeliski
"Efficient high-resolution stereo matching using local plane sweeps"
Proc. CVPR, pp. 1582-1589,
24. N. Snavely, S. M. Seitz and R. Szeliski
"Modeling the world from internet photo collections"
IJCV, vol. 80, no. 2, pp. 189-210, 2008
25. C. Strecha, W. V. Hansen, L. V. Gool, P. Fua and U. Thoen-nessen
"On benchmarking camera calibration and multi-view stereo for high resolution imagery"
Proc. CVPR, pp. 1-8,
26. J.-P. Tardif, A. Bartoli, M. Trudeau, N. Guilbert and S. Roy
"Algorithms for batch matrix factorization with application to structure-from-motion"
Proc. CVPR, pp. 1-8,
27. R. Toldo and A. Fusiello
"Robust multiple structures estimation with j-linkage"
Proc. ECCV, pp. 537-547,

28. B. Triggs

"Autocalibration from planar scenes"

Proc. ECCV, pp. 89-105,

29. B. Triggs, P. F. McLauchlan, R. I. Hartley
and A. W. Fitzgibbon

"Bundle adjustment: modern synthesis"

Vision algorithms: theory and practice, pp.
298-372, 2000

30. S. Ullman

"Maximizing rigidity: the incremental recovery
of 3-d structure from rigid and rubbery motion"

1983

31. M. Werman and A. Shashua

"The study of 3d-from-2d using elimination"

Proc. ICCV, pp. 473-479,

32. Z. Zhou, H. Jin and Y. Ma

"Robust plane-based structure from motion"

Proc. CVPR, pp. 1482-1489,

