Towards Linear-time Incremental Structure from Motion
从运动走向线性时间增量结构

Changchang Wu
吴
University of Washington
华盛顿大学

Abstract
摘要

The time complexity of incremental structure from motion (SfM) is often known as $O(n^4)$ with respect to the number of cameras. As bundle adjustment (BA) being significantly improved recently by preconditioned conjugate gradient (PCG), it is worth revisiting how fast incremen-tal SfM is. We introduce a novel BA strategy that provides good balance between speed and accuracy. Through algorithm analysis and extensive experiments, we show that incremental SfM requires only O(n) time on many ma-jor steps including BA. Our method maintains high ac-curacy by regularly re-triangulating the feature matches that initially fail to triangulate. We test our algorithm on large photo collections and long video sequences with various settings, and show that our method offers state of the art performance for large-scale reconstructions. The presented algorithm is available as part of VisualSFM at

http://homes.cs.washington.edu/˜ccwu/vsfm/.

增量结构的时间复杂度(SfM)通常被称为 o(n4)相对于摄像机的数量。随着预条件共轭梯度(PCG)技术的发展，光束法平差(BA)得到了显著的改善，因此值得重新考虑渐进式 SfM 的速度有多快。我们介绍了一种新的 BA 策略，提供了良好的平衡之间的速度和准确性。通过算法分析和大量实验，我们发现增量 SfM 在包括 BA 在内的多个步骤上只需要 o(n)时间。我们的方法保持高精度，通过定期重新三角化的特征匹配，最初没有三角化。我们在不同设置的大型照片集和长视频序列上测试了我们的算法，并表明我们的方法为大规模重建提供了最先进的性能。本文提出的算法可作为 VisualSFM 的一部分在 http://homes.cs.washington.edu/ccwu/vsfm/。

1. Introduction
1.引言

Structure from motion (SfM) has been successfully used for the reconstruction of increasingly large uncontrolled photo collections [8, 2, 5, 4, 7]. Although a large photo collection can be reduced to a subset of iconic images [8, 5] or skeletal images [15, 2] for reconstruction, the commonly known $O(n^4)$ cost of the incremental SfM is still high for large scene components. Recently an $O(n^3)$ method was demonstrated by using a combination of discrete and con-tinuous optimization [4]. Nevertheless, it remains crucial for SfM to achieve low complexity and high scalability.

运动结构(SfM)已经成功地用于重建越来越大的不受控制的照片集[8,2,5,4,7]。虽然一个大的照片集可以减少到一个子集的标志性图像[8,5]或骨架图像[15,2]的重建，通常所知的 o(n4)的增量 SfM 成本仍然是高的大场景组件。最近，利用离散优化和连续优化相结合的方法证明了一种 o(n3)方法[4]。尽管如此，SfM 实现低复杂度和高可伸缩性仍然至关重要。

This paper demonstrates our efforts for further improving the efficiency of SfM by the following contributions:

本文件展示了我们为进一步提高可持续森林管理的效率所作的努力，其贡献如下:

We introduce a preemptive feature matching that can reduce the pairs of image matching by up to 95% while still recovering sufficient good matches for reconstruction.
我们提出了一种先发制人的特征匹配方法，在恢复足够的匹配重建图像的同时，可以使图像匹配的对数减少 95%以上。

We analyze the time complexity of the conjugate gradient bundle adjustment methods. Through theoretical analysis and experimental validation, we show that the bundle adjustment requires O(n) time in practice.
我们分析了共轭梯度光束法平差方法的时间复杂度。通过理论分析和实验验证，我们表明，光束法平差需要 o(n)时间在实践中。

We show that many sub-steps of incremental SfM,
我们展示了增量 SfM 的许多子步骤,

including BA and point filtering, require only O(n) time in practice when using a novel BA strategy.
包括 BA 和点滤波，在实际应用中使用新的 BA 策略只需要 o(n)时间。
Without sacrificing the time-complexity, we introduce a re-triangulation step to deal with the problem of accumulated drifts without explicit loop closing.
在不牺牲时间复杂度的前提下，引入了一种重三角化方法来处理无显式闭环的累积漂移问题。

The rest of the paper is organized as follows: Section 2 gives the background and related work. We first introduce a preemptive feature matching in Section 3. We analyze the time complexity of bundle adjustment in Section 4 and propose our new SfM algorithms in Section 5. The experiments and conclusions are given in Section 6 and Section 7.

论文的其余部分安排如下:第二部分介绍了背景和相关工作。我们首先在第 3 部分中介绍了一种抢占式特征匹配。我们在第 4 部分分析了光束法平差的时间复杂度，并在第 5 部分提出了新的 SfM 算法。实验和结论分别在第 6 节和第 7 节中给出。

2. Related Work
2.相关工作

In a typical incremental SfM system (e.g. [14]), two-view reconstructions are first estimated upon successful feature matching between two images, 3D models are then reconstructed by initializing from good two-view reconstructions, repeatedly adding matched images, triangulating feature matches, and bundle-adjusting the structure and motion. The time complexity of such an incremental SfM algorithms is commonly known to be $O(n^4)$ for n images, and this high computation cost impedes the application of such a simple incremental SfM on large photo collections.

在一个典型的增量式 SfM 系统(例如[14])中，首先通过两幅图像之间的特征匹配估计出双视图重构，然后通过良好的双视图重构初始化、重复添加匹配图像、三角匹配、捆绑调整结构和运动来重构三维模型。这种增量 SfM 算法的时间复杂度通常被认为是 n 幅图像的 o(n4)，这种高计算量阻碍了这种简单的增量 SfM 在大型图片集上的应用。

Large-scale data often contains lots of redundant information, which allows many computations of 3D recon-struction to be approximated in favor of speed, for example:

大规模数据往往包含大量冗余信息，这使得许多 3D 重建的计算可以近似为有利于速度的方式，例如:

Image Matching Instead of matching all the image to each other, Agarwal et al. [2] first identify a small number of candidates for each images by using the vocabulary tree recognition [11], and then match the features by using approximate nearest neighbor. It is shown that the two-folded approximation of image matching still preserves enough feature matches for SfM. Frahm et al. [5] exploit the approximate GPS tags and match images only to the nearby ones. In this paper, we present a preemptive feature matching to further improve the matching speed.
图像匹配 Agarwal 等人[2]首先使用词汇树识别方法为每个图像确定少量候选图像，然后使用近似最近邻匹配特征。结果表明，图像匹配的双折叠近似仍然保留了足够的特征匹配。弗拉姆等人[5]利用近似的 GPS 标签和匹配图像只有附近的。为了进一步提高匹配速度，本文提出了一种抢占式特征匹配算法。

Bundle Adjustment Since BA already exploits linear approximations of the non-linear optimization problem, it is often unnecessary to solve the exact descent steps. Recent algorithms have achieved significant speedup by using Preconditioned Conjugate Gradient (PCG) to approximately
由于 BA 已经利用了非线性光束法平差最佳化问题的线性近似，所以通常没有必要求解准确的下降步骤。近年来，利用预条件共轭梯度(PCG)近似加速算法取得了显著的加速效果

solve the linear systems [2, 3, 1, 16]. Similarly, there is no need to run full BA for every new image or always wait until BA/PCG converges to very small residuals. In this paper, we show that linear time is required by bundle adjustments for large-scale uncontrolled photo collections.
求解线性方程组[2,3,1,16]。类似地，没有必要为每个新映像运行完整的 BA，或者总是等待 BA/pcg 汇聚到非常小的残差。在本文中，我们表明，线性时间是需要捆绑调整的大规模不受控制的照片收藏。

Scene Graph Large photo collections often contain more than enough images for high-quality reconstructions. The efficiency can be improved by reducing the number of images for the high-cost SfM. [8, 5] use the iconic images as the main skeleton of scene graphs, while [15, 2] extract skeletal graphs from the dense scene graphs. In practice, other types of improvements to large-scale SfM should be used jointly with scene graph simplifications. However, to push the limits of incremental SfM, we consider the reconstruction of a single connected component without simplifying the scene graphs.
场景图大型图片集通常包含足够的图片用于高质量的重建。对于高成本的 SfM，可以通过减少图像数量来提高效率。[8,5]使用标志性图像作为场景图的主骨架，而[15,2]从稠密的场景图中提取骨架图。在实践中，对大规模 SfM 的其他类型的改进应该与场景图简化相结合使用。然而，为了突破增量式 SfM 的极限，我们考虑在不简化场景图的情况下重建单个连接元件(图论)。

In contrast to incremental SfM, other work tries to avoid the greedy manner. Gherard et al. [6] proposed a hierarchical SfM through balanced branching and merging, which lowers the time complexity by requiring fewer BAs of large models. Sinha et al. [13] recover the relative camera rotations from vanishing points, and converts SfM to efficient 3D model merging. Recently, Crandall et al. [4] exploit GPS-based initialization and model SfM as a global MRF optimization. This work is a re-investigation of incremental SfM, and still shares some of its limitations, such as initializations affecting completeness, but does not rely on additional calibrations, GPS tags or vanishing point detection.
与渐进式 SfM 相反，其他工作试图避免贪婪的方式。Gherard 等人[6]提出了一种通过平衡分支和合并的层次结构 SfM，它通过减少大型模型的基本数据量来降低时间复杂度。Sinha 等人[13]恢复相关的摄像机旋转从消失点，并转换为高效的三维模型合并 SfM。最近，Crandall 等人利用基于 gps 的初始化和模型 SfM 作为全局 MRF 优化。这项工作是对增量 SfM 的重新研究，仍然具有一些局限性，例如影响完整性的初始化，但不依赖于附加校准、GPS 标签或消失点检测。

Notations We use n, p and q to respectively denote the number of cameras, points and observations during a reconstruction. Given that each image has a limited number of features, we have p = O(n) and q = O(n). Since this paper considers a single connected scene graph, n is also used as the number of input images with abuse of notation.
符号我们用 n，p 和 q 分别表示重建过程中摄像机、点和观测点的数量。考虑到每个图像的特征数量有限，我们有 po(n)和 qo(n)。由于本文考虑的是一个单连通的场景图，因此 n 也被用作具有符号滥用的输入图像的数目。

3. Preemptive Feature Matching
3.抢占式特征匹配

Image matching is one of the most time-consuming steps of SfM. The full pairwise matching takes $O(n^2)$ time for n input images, however it is fine for large datasets to compute a subset of matching (e.g. by using vocabulary tree [2]), and the overall computation can be reduced to O(n). In addition, image matching can be easily paral-lelized onto multiple machines [2] or multiple GPUs [5] for more speedup. Nevertheless, feature matching is still one of the bottlenecks because typical images contain several thousands of features, which is the aspect we try to handle.
图像匹配是 SfM 中最耗时的步骤之一。对于 n 个输入图像，完全成对匹配需要 o(n2)时间，但是对于大型数据集计算匹配的子集(例如使用词汇树[2])是可行的，并且整体计算可以减少到 o(n)。此外，图像匹配可以很容易地在多台计算机[2]或多个 gpu[5]上进行部分匹配，以获得更高的加速比。然而，特征匹配仍然是一个瓶颈，因为典型的图像包含几千个特征，这是我们试图处理的方面。

Due to the diversity of viewpoints in large photo collections, the majority of image pairs do not match (75% 98% for the large datasets we experimented). A large portion of matching time can be saved if we can identify the good pairs robustly and efficiently. By exploiting the
由于大型照片采集中视点的多样性，大多数图像对不匹配(对于我们实验的大型数据集，匹配率为 75%98%)。如果能够稳定有效地识别好的匹配对，可以节省大量的匹配时间。通过利用

scales of invariant features [10], we propose the following preemptive feature matching for this purpose:
为此，我们提出了以下优先的特征匹配方法:

1. Sort the features of each image into decreasing scale order. This is a one-time O(n) preprocessing.
将每幅图像的特征按缩放顺序排序，这是一次性的 o(n) 预处理。
2. Generate the list of pairs that need to be matched, either using all the pairs or a select a subset ([2, 5]).
生成需要匹配的对的列表，要么使用所有对，要么选择一个子集([2,5])。
3. For each image pair (parallelly), do the following:
对于每个图像对(并行)，执行以下操作:
   (a) Match the first h features of the two images.
   匹配两个图像的第一个 h 特征。
   (b) If the number of matches from the subset is smaller than t_h, return and skip the next step.
   如果子集中匹配的数量小于，返回并跳过下一步。
   (c) Do regular matching and geometry estimation.
   进行规则匹配和几何估计。

where h is the parameter for the subset size, and t_h is the threshold for the expected number of matches. The feature matching of subset and full-set uses the same nearest neighbor algorithm with distance ratio test [10] and require the matched features to be mutually nearest.
其中 h 是子集大小的参数，th 是预期匹配数量的阈值。子集和全集的特征匹配采用距离比检验[10]相同的最近邻算法，并要求匹配的特征是相互最近的。

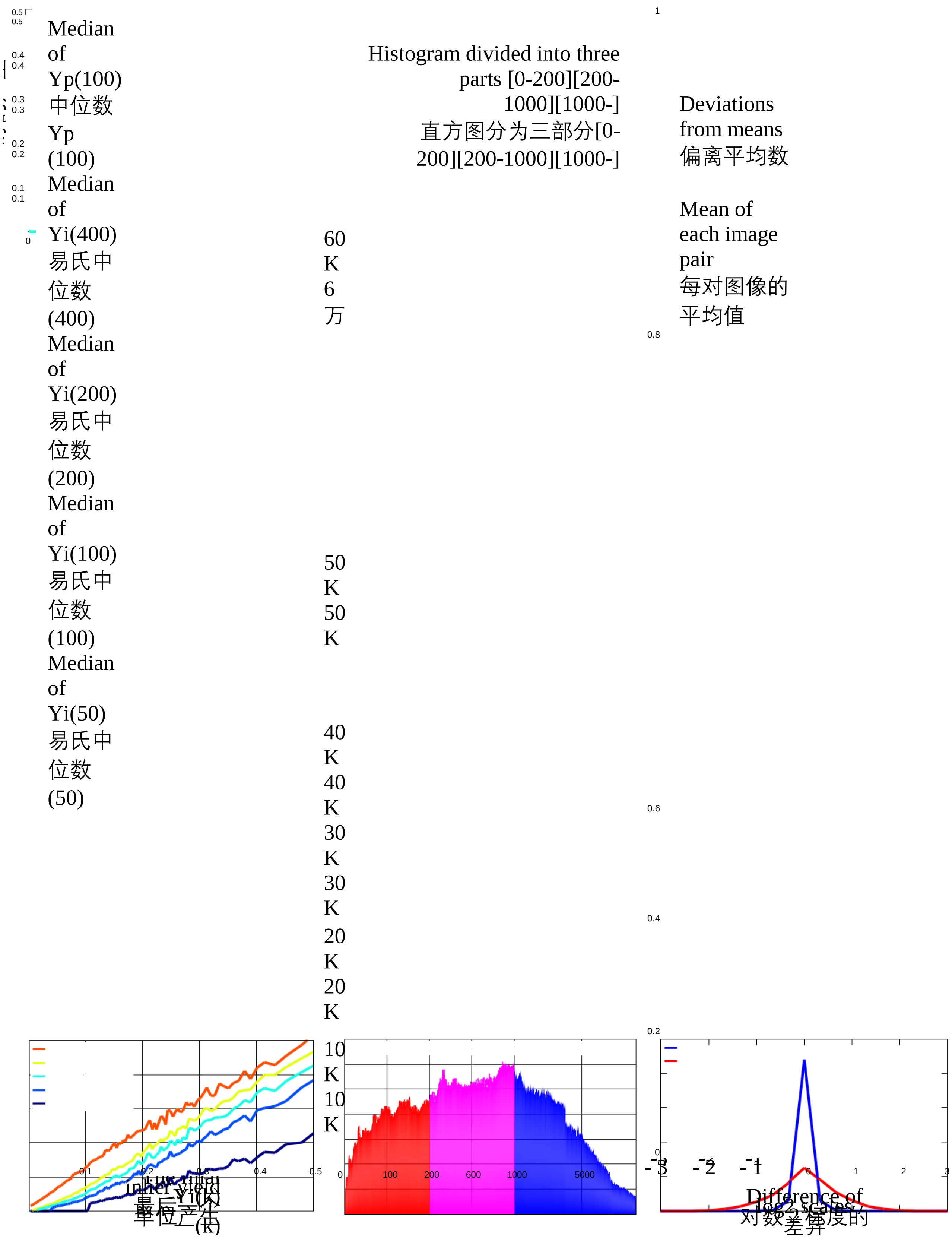
Let k_1 and k_2 be the number of features of two images, and $k = \max(k_1; k_2)$. Let $m_p(h)$ and $m_i(h)$ be the number of putative and inlier matches when using up to h top-scale features. We define the yield of feature matching as
设 k1 和 k2 为两个图像的特征数，kmax(k1;k2)。设 mp(h) 和 mi(h)为假定匹配和内点匹配的数目，当使用最多个顶级特性时。我们将特征匹配的得率定义为

$$Y_p(h) = \frac{m_p(h)}{h} \text{ and } Y_i(h) = \frac{m_i(h)}{h} \quad (1)$$

$$Y_p(h) = \frac{m_p(h)}{h} \text{ 及易}(h) = \frac{M_i(h)}{h} : \quad (1)$$

We are interested in how the yields of the feature subset cor-relate with the final yield $Y_i(k)$. As shown in Figure 1(a), the distributions of $Y_p(h)$ and $Y_i(h)$ are closely related to $Y_i(k)$ even for very small h such as 100. Figure 1(b) shows that the chances of have a match within the top-scale features are on par with other features. This means that the top-scale subset has roughly h=k chance to preserve a match, which is much higher than the $h^2=(k_1k_2)$ chance of random sampling h features. Additionally, the matching time for the majority is reduced to a factor of $h^2=(k_1k_2)$ along with better caching. In this paper, we choose h = 100 with consideration of both efficiency and robustness.
我们感兴趣的是特征子集的产量与最终产量的关系。如图 1(a)所示，Yp(h)和 Yi(h)的分布与 Yi(k)密切相关，即使在非常小的 h，如 100。图 1(b)显示了在顶级特性中匹配的几率与其他特性相当。这意味着顶级子集保持匹配的几率大约为 hk，远高于随机抽样 h 特征的 h2(k1k2)几率。此外，大多数的匹配时间减少到 h2(k1k2)因子，同时有更好的缓存。在本文中，我们选择 h100，同时考虑了效率和鲁棒性。

The top-scale features match well for several reasons: 1) A small number of top-scale feature can cover a relatively large scale range due to the decreasing number of features on higher Gaussian levels. 2) Feature matching is well structured such that the large-scale features in one image often match with the large-scale features in another. The scale variation between two matched features is jointly determined by the camera motion and the scene structure, so the scale variations of multiple feature matches are not independent to each other. Figure 1(c) shows our statistics of the feature scale variations. The feature scale variations of an image pair usually have a small variance due to small viewpoint changes or well-structured scene depths. For the same reasons, we use up to 8192 features for the regular feature matching, which are sufficient for most scenarios. While large photo collections contain redundant views and features, our preemptive feature matching allows putting most efforts in the ones that are more likely to be matched.
顶层特征匹配的好处有以下几个原因:1)由于高斯层次上特征数量的减少，少量的顶层特征可以覆盖相对较大的尺度范围。2)特征匹配结构良好，一幅图像中的大规模特征往往与另一幅图像中的大规模特征匹配。两个匹配特征之间的尺度变化是由摄像机运动和场景结构共同决定的，因此多个特征匹配的尺度变化并不是相互独立的。图 1(c)显示了我们对特征尺度变化的统计。图像对的特征尺度变化通常由于视点变化小或场景深度结构良好而变化很小。出于同样的原因，我们使用多达 8192 个特性进行常规特性匹配，这对于大多数场景已经足够了。虽然大型图片集包含多余的视图和特征，但是我们优先的特征匹配允许将大部分精力放在更有可能匹配的特征上。



(a) Yield when using subset of features (b) Histogram of max index of a feature match (c) The distribution of feature scale variations

(a)使用特征子集时的产量(b)特征匹配的最大指数直方图(c)特征尺度变化的分布

Figure 1. (a) shows the relationship between the final yield and the yield from a subset of top-scale features. For each set of image pairs that have roughly the same final yield, we compute the median of their subset yield for different h. (b) gives the histogram of the max of two indices of a feature match, where the index is the position in the scale-decreasing order. We can see that the chances of matching within

top-scale features are similar to other features. (c) shows two distributions of the scale variations computed from 130M feature matches. The mean scale variations between two images are given by the red curve, while the deviations of scale variation from means are given by the blue one. The variances of scale changes are often small due to small viewpoint changes or structured scene depths.

图 1。(a)显示最终产量与来自一组顶层特征的产量之间的关系。对于具有大致相同最终产量的每组图像对，我们计算不同时间的子集产量的中位数(b)给出了特征匹配的两个指标的最大值的直方图，其中指标是按比例递减顺序排列的位置。我们可以看到，在顶级特征中匹配的几率与其他特征相似。(c)显示了 130M 特征匹配计算的两种尺度变化分布。两幅图像之间的平均尺度变化用红曲线表示，尺度变化与平均值的偏差用蓝曲线表示。尺度变化的差异往往是很小的，由于小的观点变化或结构化的场景深度。

4. How Fast Is Bundle Adjustment?
4.光束法平差有多快？

Bundle adjustment (BA) is the joint non-linear optimization of structure and motion parameters, for which Levenberg-Marquardt (LM) is method of choice. Recently, the performance of large-scale bundle adjustment has been significantly improved by Preconditioned Conjugate Gradient (PCG) [1, 3, 16], hence it is worth re-examining the time complexity of BA and SfM.

光束法平差是结构和运动参数的联合非线性优化，对此 Levenberg-Marquardt(LM)是首选方法。最近，预条件共轭梯度(PCG)[1,3,16]已经显著地改善了大规模光束法平差的性能，因此，BA 和 SfM 的时间复杂度值得重新研究。

Let x be a vector of parameters and f(x) be the vector of reprojection errors for a 3D reconstruction. The optimization we wish to solve is the non-linear least squares problem: x = arg min$_x$ kf(x)k^2. Let J be the Jacobian of f(x) and a non-negative diagonal vector for regulariza-tion, then LM repeatedly solves the following linear system

设 x 为参数矢量，f(x)为三维重建中重投影误差矢量。我们要解决的最优化问题是非线性最小二乘问题:xargminxkf(x)k2。设 j 为 f(x)的雅可比矩阵和正则化的非负对角向量，然后 LM 重复求解下列线性系统

$$(J^T J + \ ) \qquad = \qquad J^T f;$$

(JT j +)

and updates x x + if kf(x +)k < kf(x)k. The matrix H = J^T J + is known as the augmented Hessian matrix. Gaussian elimination is typically used to first obtain a reduced linear system of the camera parameters, which is called Schur Complement.

并更新 xx+ifkf(x+)kkf(x)k。矩阵 hJTj+称为增广黑森矩阵。高斯消去法通常用于首先获得一个简化的摄像机参数线性系统，这个系统被称为舒尔补。

The Hessian matrix require O(q) = O(n) space and time to compute, while Schur complement requires O(n^2) space and time to compute. It takes cubic time O(n^3) or O(p^3) to solve the linear system by Cholesky factorization. Because the number of cameras is much smaller than the number of points, the Schur complement method can reduce the factorization time by a large constant factor. One impressive example of this algorithm is Lourakis and Argyros' SBA [9] used by Photo Tourism [14].

黑森矩阵需要 o(q)o(n)空间和时间来计算，而舒尔补则需要 o(n2)空间和时间来计算。用 Cholesky 分解法求解线性系统需要三次时间 o(n3)或 o(p3)。因为摄像机的数量比点的数量少得多，舒尔补方法可以减少因子分解的时间一个大的常数因子。这种算法的一个令人印象深刻的例子是 Lourakis 和 Argyros 的 SBA[9]，由 PhotoTourism[14]使用。

For conjugate gradient methods, the dominant computation is the matrix-vector multiplications in multiple CG iteration, of which the time complexity is determined by the size of the involved matrices. By using only the O(n) space Hessian matrices and avoiding the Schur complement, the

对于共轭梯度法，主要的计算方法是多次 CG 迭代中的矩阵-矢量乘法，其时间复杂度取决于所涉及矩阵的大小。通过仅使用 o(n)空间 Hessian 矩阵和避免舒尔补，

CG iteration has achieved O(n) time complexity [1, 3]. Recently, the multicore bundle adjustment [16] takes one step further by using implicit multiplication of the Hessian matrices and Schur Complements, which requires to construct only the O(n) space Jacobian matrices. In this work, we use the GPU-version of multicore bundle adjustment. Figure 2(a) shows the timing of CG iterations from our bundle adjustment problems, which exhibits linear

Cg 迭代实现了 o(n)时间复杂度[1,3]。最近，多核光束法平差通过使用 Hessian 矩阵和 Schur 补数的隐式乘法更进一步，这只需要构造 o(n)空间 Jacobian 矩阵。在这项工作中，我们使用多核光束法平差的 gpu 版本。图 2(a)显示了从我们的光束法平差问题到 CG 迭代的时间，这是线性的

relationship between the time T_{cg} of a CG iteration and n. p

一个 CG 迭代的 Tcg 时间和 n.p 之间的关系

In each LM step, PCG requires $O(\sqrt{})$ iterations to accurately solve a linear system [12], where is the condition number of the linear system. Small condition numbers can be obtained by using good pre-conditioners, for example, quick convergence rate has been demonstrated with block-jacobi preconditioner [1, 16]. In our experiments, PCG uses an average of 20 iterations solve a linear system.

在每个 LM 步骤中，PCG 需要 o()迭代来快速求解一个线性系统[12]，其中的条件数是线性系统的条件数。使用良好的预处理器可以得到较小的条件数，例如，块-jacobi 预处理器[1,16]证明了收敛速度很快。在我们的实验中，PCG 平均用 20 次迭代来解决一个线性系统。

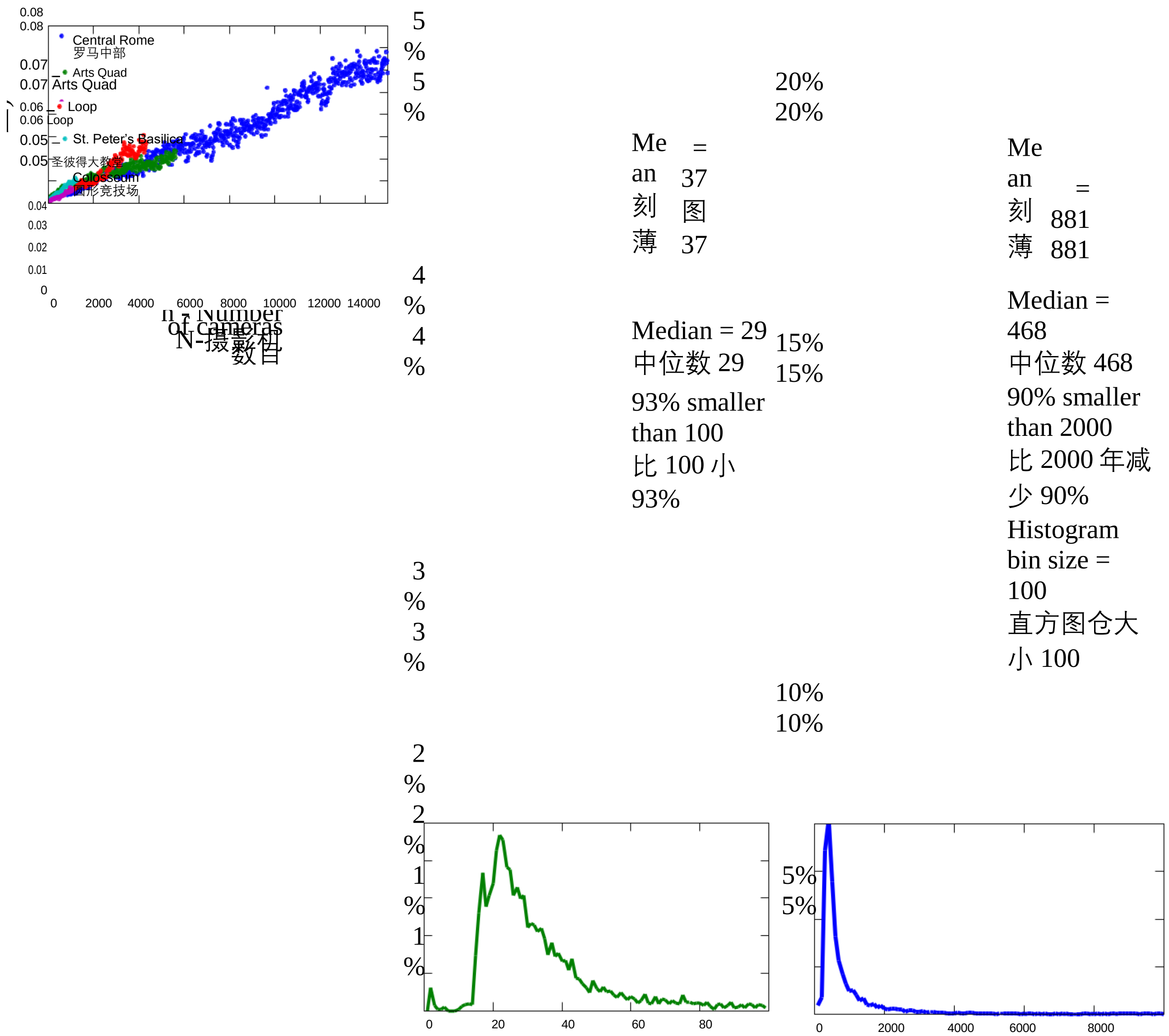
Surprisingly, the time complexity of bundle adjustment has already reached O(n), provided that there are O(1) CG iterations in a BA. Although the actual number of the CG/LM iterations depends on the difficulty of the input problems, the O(1) assumption for CG/LM iterations is well supported by the statistics we collect from a large number of BAs of varying problem sizes. Figure 2(b) gives the distribution of LM iterations used by a BA, where the average is 37 and 93% BAs converge within less than 100 LM iterations. Figure 2(c) shows the distribution of the total CG iterations (used by all LM steps of a BA), which are also normally small. In practice, we choose to use at most 100 LM iterations per BA and at most 100 CG iterations per LM, which guarantees good convergence for most problems.

令人惊讶的是，在 BA 中有 o(1)CG 迭代的情况下，光束法平差的时间复杂度已经达到 o(n)。虽然 cg/lm 迭代的实际数量取决于输入问题的难度，但是我们从大量问题大小不同的 BAs 中收集的统计数据很好地支持了 cg/lm 迭代的 o(1)假设。图 2(b)给出了 BA 使用的 LM 迭代的分布，其中平均值为 37%和 93%的 BAs 在少于 100 个 LM 迭代中收敛。图 2(c)显示了 CG 迭代总数的分布(BA 的所有 LM 步骤都使用)，这些迭代通常也很小。在实践中，我们选择每 BA 最多使用 100 个 LM 迭代，每 LM 最多使用 100 个 CG 迭代，这保证了大多数问题的良好收敛性。

5. Incremental Structure from Motion
5.运动增量结构

This section presents the design of our SfM that practically has a linear run time. The fact that bundle adjustment can be done in O(n) time opens an opportunity of push-

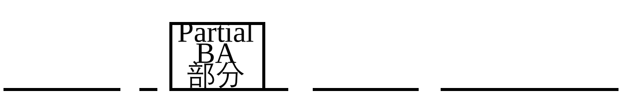
本节介绍了实际上具有线性运行时间的 SfM 的设计。光束法平差可以在 o(n)时间内完成，这一事实提供了一个推-的机会



(a) Time spent on a single CG iteration. (b) Number of LM iterations used by a BA (c) Number of CG iterations used by a BA
(a)花费在一次 CG 迭代上的时间。(b)BA 使用的 LM 迭代次数(c)BA 使用的 CG 迭代次数

Figure 2. Bundle adjustment statistics. (a) shows that T_{cg} is roughly linear to n regardless of the scene graph structure. (b) and (c) show the distributions of the number of LM and CG iterations used by a BA. It can be seen that BA typically converges within a small number of LM and CG iterations. Note that we consider BA converges if the mean squared errors drop to below 0:25 or cannot be decreased.
图 2。光束法平差统计。(a)表明，不管场景图结构如何，Tcg 大致上是线性的。(b)和(c)显示了 BA 使用的 LM 和 CG 迭代次数的分布。可以看出，BA 通常在少量 LM 和 CG 迭代中收敛。注意，如果均方误差下降到 0:25 以下，或者不能降低，我们考虑 BA 收敛。

ing incremental reconstruction time closer to O(n). As illustrated in Figure 3, our algorithm adds a single image at each iteration, and then runs either a full BA or a partial BA. After BA and filtering, we take another choice between continuing to the next iteration and a re-triangulation step.
增量重建时间接近 o(n)。如图 3 所示，我们的算法在每次迭代中添加一个图像，然后运行完整的 BA 或部分 BA。在 BA 和过滤之后，我们在继续下一个迭代和重新三角化步骤之间进行另一个选择。



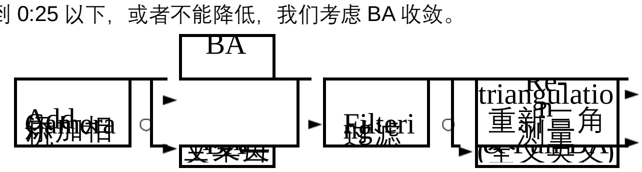


Figure 3. A single iteration of the proposed incremental SfM, which is repeated until no images can be added to reconstruction.
图 3。一个单一的迭代建议增量 SfM，这是重复，直到没有图像可以添加到重建。

5.1. How Fast Is Incremental SfM?
5.1.增量 SfM 有多快？

The cameras and 3D points normally get stabilized quickly during reconstruction, thus it is unnecessary to optimize all camera and point parameters at every iteration. A typical strategy is to perform full BAs after adding a constant number of cameras (e.g.), for which the accumulated time of all the full BAs is

在重建过程中，摄像机和三维点通常很快得到稳定，因此不需要在每次迭代中优化所有摄像机和点的参数。一个典型的策略是在添加一定数量的摄像机(例如)之后执行完整的 BAs，对于这些摄像机，所有完整的 BAs 的累计时间是

bn= c TBA(i) = O bn= c (i) = O n2 N 2 ; (2)
Bn c Tba (i) o Bn c(i) o (二)
x x

when using PCG. This is already a significant reduction from the $O(n^4)$ time of Cholesky factorization based BA. As the models grow larger and more stable, it is affordable to skip more costly BAs, so we would like to study how much further we can go without losing the accuracy.

使用 PCG 时。这已经从 o(n4)时间的 Cholesky 分解为基础的 BA 显着减少。随着模型变得越来越大、越来越稳定，跳过更昂贵的 BAs 是可以承受的，所以我们想要研究在不丧失准确性的情况下，我们还能走多远。

In this paper, we find that the accumulated time of BAs can be reduced even more to O(n) by using a geomet-ric sequence for full BAs. We propose to perform full optimizations only when the size of a model increases relatively by a certain ratio r (e.g. 5%), and the resulting time spent on full BAs approximately becomes

在本文中，我们发现使用一个完整的 ba 的风水序列可以将 ba 的累积时间减少到 o(n)。我们建议只有当模型的大小相对增加一定的比率 r(例如 5%)时才执行完全优化，并且在完全基础上花费的时间大约变成

1 TBA 总胆固醇 n = O ((1 + r)i 哦 1 n = O ((1 + r)i 哦 n : (3)
x x : (3)

Although the latter added cameras are optimized by fewer full BAs, there are normally no accuracy problems because the full BAs always improve more for the parts that have larger errors. As the model gets larger, more cameras are added before running a full BA. In order to reduce the ac-cumulated errors, we keep running local optimizations by using partial BAs on a constant number of recently added cameras (we use 20) and their associated 3D points. Such partial optimizations involve O(1) cameras and points pa-rameters, so each partial BA takes O(1) time. Therefore, the time spent on full BAs and partial BAs adds to O(n), which is experimentally validated by the time curves in Figure 6.

虽然后者增加的摄像头优化了较少的完整的 ba，通常没有精度问题，因为完整的 ba 总是改善更多的部分，有较大的误差。随着模型越来越大，在运行完整的 BA 之前会添加更多的摄像头。为了减少交流累积误差，我们继续运行局部优化使用部分 ba 对一个常数的新增摄像机(我们使用 20)及其相关的 3D 点。这种部分优化涉及到 o(1)摄像机和点探测器，因此每个部分 BA 需要 o(1)时间。因此，花在完整的和部分的 ba 上的时间增加到 o(n)，这是通过图 6 中的时间曲线实验验证的。

Following the BA step, we do filtering of the points that have large reprojection errors or small triangulation angles. The time complexity of a full point filtering step is O(n). Fortunately, we only need to process the 3D points that have been changed, so the point filtering after a partial BA can be done O(1) time. Although each point filtering after a full BA takes O(n) time, they add to only O(n) due to the geometric sequence. Therefore, the accumulated time on all point filtering is also O(n).

在 BA 步骤之后，我们对重投影误差较大或三角测量角度较小的点进行滤波。全点过滤步骤的时间复杂度为 o(n)。幸运的是，我们只需要处理已经更改的 3D 点，因此，经过部分 BA 后的点过滤可以在 o(1)时间内完成。尽管经过完整的 BA 后，每个点过滤需要 o(n)时间，但由于几何序列，它们只添加到 o(n)。因此，所有点过滤的累积时间也是 o(n)。

Another expensive step is to organize the resection candi-dates and to add new cameras to 3D models. We keep track of the potential 2D-3D correspondences during SfM by using the feature matches of newly added images. If each image is matched to O(1) images, which is a reasonable as-sumption for large-scale photo collections, it requires O(1) time to update the correspondence information at each iter-ation. Another O(1) time is needed to add the camera to the model. The accumulated time of these steps is again O(n).

另一个昂贵的步骤是组织切除烛光日期和添加新的摄像机到 3D 模型。我们利用新增图像的特征匹配来跟踪 SfM 过程中可能出现的 2D-3D 对应。如果每个图像都与 o(1)图像匹配，这对于大规模的图像集合来说是一个合理的消耗，则需要 o(1)时间来更新每个变量的通信信息。另一个 o(1)的时间是需要添加相机的模型。这些步骤的累积时间又是 o(n)。

We have shown that major steps of incremental SfM contribute to an O(n) time complexity. However, the above analysis ignored several things: 1) finding the portion of data for partial BA and partial filtering. 2) finding the subset of images that match with a single image in the resection stage. 3) comparison of cameras during the resection stage. These steps require O(n) scan time at each step and add to $O(n^2)$ time in theory. It is possible to keep track of these subsets for further reduction of time complexity. However,

我们已经展示了增量 SfM 的主要步骤对 o(n)时间复杂性的贡献。然而，上面的分析忽略了几件事情:1)找到部分 BA 和部分过滤的数据部分。2)在后方交会阶段找到与单幅图像匹配的图像子集。3)切除期间摄影机的比较。这些步骤需要 o(n)扫描时间在每个步骤和增加 o(n2)时间在理论上。为了进一步降低时间复杂度，可以跟踪这些子集。然而，

since the O(n) part dominates the reconstruction in our experiments (up to 15K, see Table 2 for details), we have not tried to further optimize these O(n^2) steps.
由于 o(n)部分在我们的实验中占主导地位(直到 15K，详见表 2)，我们没有尝试进一步优化这些 o(n2)步骤。

5.2. Re-triangulation (RT)
5.2.重新三角测量(RT)

Incremental SfM is known to have drifting problems due to the accumulated errors of relative camera poses. The constraint between two camera poses is provided by their triangulated feature matches. Because the initially estimated poses and even the poses after a partial BA may not be accurate enough, some correct feature matches may fail to triangulate for some triangulation threshold and filtering threshold. The accumulated loss of correct feature matches is one of the main reasons of the drifting.

增量式 SfM 由于相对摄像机姿态的累积误差而存在漂移问题。两个摄像机姿态之间的约束通过它们的三角特征匹配来提供。由于初始估计的姿态，甚至部分 BA 后的姿态可能不够准确，对于某些三角化阈值和滤波阈值，一些正确的特征匹配可能无法进行三角化。正确特征匹配的累积损失是引起特征漂移的主要原因之一。

To deal with this problem, we propose to re-triangulate (RT) the failed feature matches regularly (with delay) dur-ing incremental SfM. A good indication of possible bad rel-ative pose between two cameras is a low ratio between their common points and their feature matches, which we call under-reconstructed. In order to wait until the poses to get more stabilized, we re-triangulate the under-reconstructed pairs under a geometric sequence (e.g. r^0 = 25% when the size of a model increases by 25%). To obtain more points, we also increase the threshold for reprojection errors during RT. After re-triangulating the feature matches, we run full BA and point filtering to improve the reconstruction. Each RT step requires O(n) time and accumulates to the same O(n) time thanks to the geometric sequence.

为了解决这个问题，我们提出对失败的特征进行有规律的(有延迟的)重新三角测量。两个摄像机之间可能存在的相对位姿的一个很好的指示是它们的共同点与它们的特征匹配之间的低比值，我们称之为欠重建。为了等待姿态变得更加稳定，我们在一个几何序列下重新对未重建的姿态对进行三角测量(例如，当模型尺寸增加 25%时，重建的姿态对为 r025%)。为了得到更多的点，我们还提高了反投影误差的阈值。对特征匹配重新进行三角剖分后，运行全 BA 和点滤波以改善重建效果。每个 RT 步骤需要 o(n)时间，并且由于几何顺序累积到相同的 o(n)时间。

The proposed RT step is similar to loop-closing, which however deals with drifts only when loops are detected. By looking for the under-reconstructed pairs, our method is able to reduce the drift errors without explicit loop detec-tions, given that there are sufficient feature matches. In fact, the RT step is more general because it works for the relative pose between any matched images, and it also makes loop detection even easier. Figure 4 shows our incremental SfM with RT on a 4K image loop, which correctly reconstruct the long loop using RT without explicit loop closing.

提出的 RT 步骤类似于闭环，但是它只处理检测到循环时的漂移。在特征匹配足够的情况下，通过寻找欠重构对，可以在不需要显式循环测试的情况下减少漂移误差。实际上，RT 步骤更加通用，因为它适用于任何匹配图像之间的相对位姿，而且它还使得循环检测更加容易。图 4 显示了我们在 4K 图像循环上使用 RT 的增量 SfM，它在没有显式循环关闭的情况下使用 RT 正确地重建了长循环。

6. Experiments
6.实验

We apply our algorithms on five datasets of different sizes. The Central Rome dataset and the Arts Quad datasets are obtained from the authors of [4], which contain 32768 images and 6514 images respectively. The Loop dataset are 4342 frames of high resolution video sequences of a street block. The St. Peter's Basilica and Colosseum datasets have 1275 and 1164 images respectively. We run all the reconstructions on a PC with an Intel Xenon 5680 3.33Ghz CPU (24 cores), 12GB RAM, and an nVidia GTX 480 GPU.

我们将我们的算法应用于五个不同大小的数据集。中央罗马数据集和艺术四方数据集获得的作者[4]，其中包含 32768 图像和 6514 图像分别。循环数据集是一个街区的 4342 帧高分辨率视频序列。圣彼得大教堂和罗马圆形大剧场的数据集分别有 1275 张和 1164 张图片。我们在一台 PC 上运行所有的重建工作，这台 PC 配有 IntelXenon56803.33GhzCPU(24 核)、12gb 内存和 nVidiaGTX480GPU。

6.1. Feature Matching
6.1.特征匹配

To allow experiment on the largest possible model, we have tried to first match sufficient image pairs. The

为了允许在最大可能的模型上进行实验，我们首先尝试匹配足够的图像对

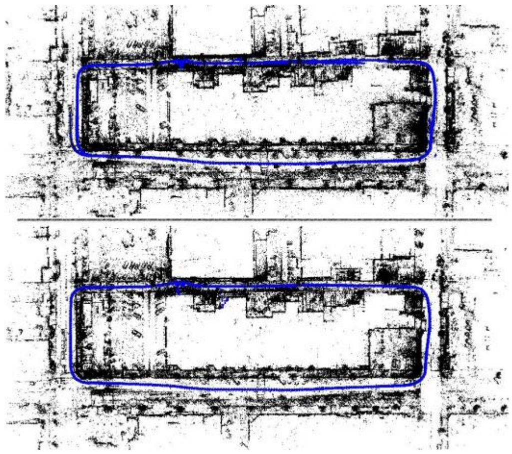


Figure 4. Our reconstruction of the Loop dataset of 4342 frames (the blue dots are the camera positions). Our algorithm correctly handles the drifting problem in this long loop by using RT.
图 4。我们重建了 4342 帧的 Loop 数据集(蓝点代表摄像机的位置)。我们的算法通过使用 rt 正确处理了这个长循环中的漂移问题。

full pair-wise matching is computed for the St. Peter's Basilica and Colosseum dataset. For the Arts Quad and Loop datasets, each image is matched to the nearby ones according to GPS. For the Central Rome dataset, our preemptive matching with $h = 100$ and $t_h = 4$ is applied.
对圣彼得大教堂和罗马圆形大剧场的数据集进行全对匹配。对于 Arts 四边形和 Loop 数据集，根据 GPS，每个图像都与附近的图像匹配。对于中央罗马数据集，我们先发制人的匹配 h100 和 th4 是应用。

We then run our incremental SfM with the subset of image matches that satisfy the preemptive matching for = 100 and different th. Table 1 shows the statistics of the feature matches and the reconstructed number of cameras of the largest reconstructed models. With the preemptive matching, we are able to reconstruct the larges SfM model of 15065 cameras for Rome and complete models for other datasets. Preemptive matching is able to significantly re-duce the number of image pairs for regular matching and still preserve a large portion of the correct feature matches. For example, 43% feature matches are obtained with 6% of image pairs for the St. Peter's. It is worth noting that it is ex-tremely fast to match the top 100 features. Our system has an average speed of 73K pairs per second with 24 threads.

然后，我们运行增量 SfM 与图像匹配的子集，满足抢占式匹配为 100 和不同的时间。表 1 显示了特征匹配的统计数据和最大重建模型的相机重建数量。通过先发制人的匹配，我们能够为罗马重建 15065 台相机的大型 SfM 模型，并为其他数据集重建完整的模型。先发制人的匹配能够显著减少图像对的数量进行常规匹配，同时仍然保留了大部分正确的特征匹配。例如，对于圣彼得教堂，43%的特征匹配是用 6%的图像对获得的。值得注意的是，匹配前 100 个功能的速度非常快。我们的系统平均速度为每秒 73K 对，有 24 个线程。

The preemptive matching has a small chance of losing weak links when using a large threshold. Complete models are reconstructed for all the datasets when $t_h = 2$. However, for $t_h = 4$, we find that one building in Arts Quad is missing due to occlusions, and the Colosseum model breaks into the interior and the exterior. We believe a more adaptive scheme of preemptive matching should be explored in the future.

当使用较大的阈值时，抢占式匹配丢失弱链接的几率很小。2 时，为所有数据集重建完整的模型。然而，对于第四，我们发现在艺术广场的一个建筑是缺失的，由于闭塞，和斗兽场模型打破了内部和外部。我们认为未来应该探索一种更具适应性的抢先匹配方案。

6.2. Incremental SfM
6.2.渐进式 SfM

We run our algorithm with all computed feature matches for the five datasets using the same settings. Table 2 summa-rizes the statistics and time of the experiment for r = 5% and r^0 = 25%. Figure 4, 5 and 7 show the screenshots of our SfM models. Our method efficiently reconstructs large, accurate and complete models with high point density. In particular, the two 1K image datasets are reconstructed in less than 10

我们使用相同的设置对五个数据集的所有计算特征匹配运行我们的算法。表 2 总结了 r5%和 r025%实验的统计数据和时间。图 4、5 和 7 显示了 SfM 模型的屏幕截图。我们的方法有效地重建了大型，准确和完整的模型与高点密度。特别是，这两个 1K 图像数据集的重建时间少于 10

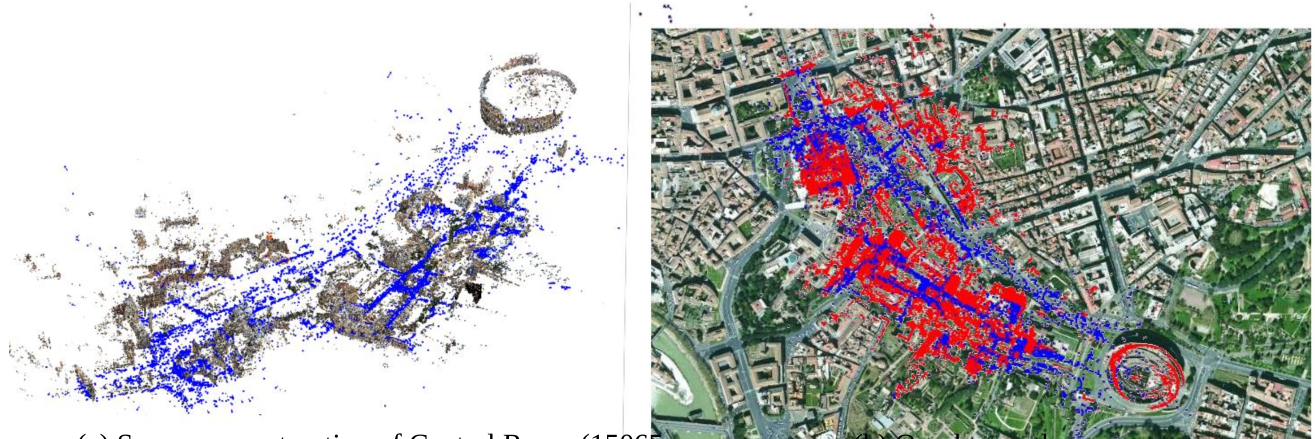
	Without Preemptive Matching 没有先发制人的匹配				Using Preemptive Matching (h = 100) 使用抢占式匹配(h 100)				
Dataset 数据集	Pairs to 对到 Match 匹配	Pairs With 配对 15+ Inliers 15 + Inliers	Feature 功能 Matches 火柴	n	th 第十二章	Pairs to 对到 Match 匹配	Pairs With 配对 15+ Inliers 15 + Inliers	Feature 功能 Matches 火柴	n
Central Rome 罗马中部	N/A 不适用	N/A 不适用	N/A 不适用	N/A 不适用	4	13551K 13551K	540K 540K	67M 67M	15065
Arts Quad 艺术方阵	15402K 15402K	192K 192K	32M 32M	5624	4	521K, 3% 521K，3%	62K, 32% 62K，32%	25M, 78% 2500 万，78%	4272
					2	4308K, 28% 4308K，28%	121K, 63% 121K，63%	29M, 91% 2900 万，91%	5393
Loop 环路	709K 709K	329K 329K	158M 158M	4342	4	269K, 38% 269K，38%	235K, 71% 235K，71%	150M, 95% 1.5 亿，95%	4342
					8	151K, 21% 151K，21%	150K, 46% 150K，46%	135M, 85% 1.35 亿美元，85%	4342
St. Peter's 圣彼得教堂	812K 812K	217K 217K	21M 21M	1267	4	46K, 6% 46K，6%	38K, 18% 38K，18%	9.1M, 43% 9.1 m，43%	1211
					8	220K, 27% 22 万，27%	100K, 46% 100K，46%	14M, 67% 1400 万，67%	1262
Colosseum 圆形竞技场	677K 677K	54K 54K	6.8M 6.8 m	1157	4	23K, 3% 23K、，3%	13K, 24% 13K，24%	4.1M, 60% 4.1 m，60%	517+426 517 + 426
					2	149K, 22% 149K，22%	28K, 52% 28K，52%	5.4M, 79% 5.4 m，79%	1071

Table 1. Reconstruction comparison for different feature matching. We first try $t_h = 4$, and then try $t_h = 8$ of the result is complete, or try $t_h = 2$ if the resulting reconstruction is incomplete. Comparably complete models are reconstructed when using preemptive matching, where only a small set of image pairs need to be matched. All reconstructions use the same setting r = 5% and r^0 = 25%.
表一。不同特征匹配的重建比较。我们首先尝试第 4 个，然后尝试第 8 个结果是完整的，或尝试第 2 个，如果结果重建是不完整的。采用抢占式匹配方法重建比较完整的模型，只需匹配少量的图像对。所有的重建都使用相同的设置 r5%和 r025%。

Dataset 数据集	Input 输入 Images 图片	Cameras 相机 n	Points 积分 p	Observa- 天文台 tions q 问题 q	Time t 时间 t Overall 整体而言	Time 时间 Full BA 全文英文字母	Time 时间 Partial BA 部分 BA	Time 时间 Adding 加入	Time 时间 Filtering 过滤
Central Rome 罗马中部	32768	15065	1660415	12903348	6010	2008	2957	549	247
Arts Quad 艺术方阵	6514	5624	819292	5838784	2132	1042	807	122	57
Loop 环路	4342	4342	1101515	7195960	3251	1731	478	523	47
St. Peter's	1275	1267	292379	2706250	583	223	268	48	20

圣彼得教堂									
Colosseum 圆形竞技场	1164	1157	293724	1759136	591	453	100	19	9

Table 2. Reconstruction summary and timing (in seconds). Only the reconstruction of largest model is reported for each dataset. The reported time of full BA includes the BAs after the RT steps. The “adding” part includes the time on updating resection candidates, estimating camera poses, and adding new cameras and points to a 3D model.
表二。重建摘要和计时(秒)。每个数据集只报告最大模型的重建。报告的完全 BA 时间包括 RT 步骤之后的 BA。“添加”部分包括更新后方交会候选方案、估计摄像机位置、添加新摄像机和点到 3D 模型的时间。



(a) Sparse reconstruction of Central Rome (15065 cameras)
(a)罗马市中心的稀疏重建(15065 架照相机)

(b) Overlay on the aerial image
(b)航空影像重叠

Figure 5. Our Rome reconstruction (the blue dots are the camera positions). The recovered structure is accurate compared to the map.
图 5。我们的罗马重建(蓝色的点是相机的位置)。与地图比较，恢复的结构是准确的。

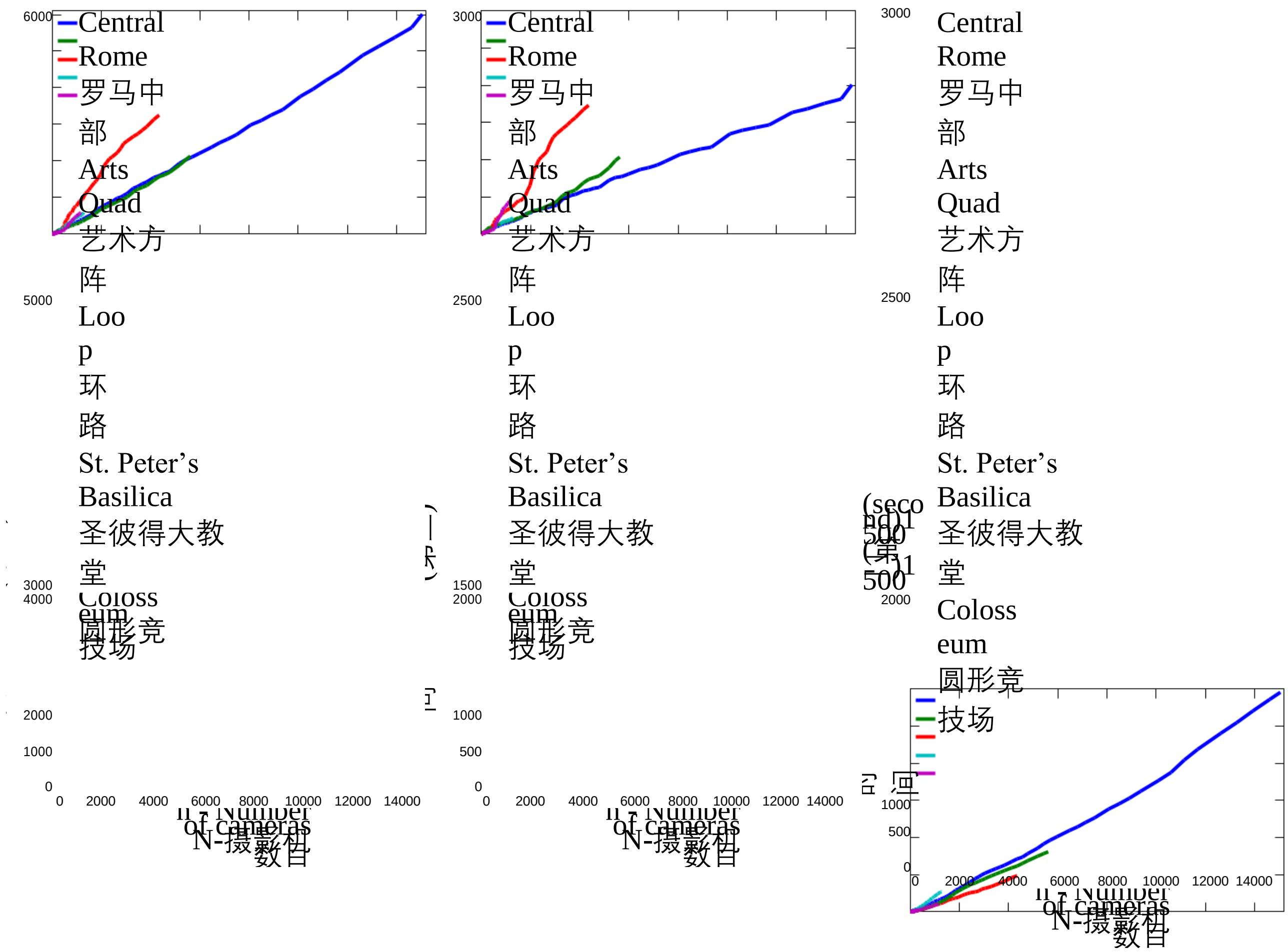
minutes, and the 15K camera model of Rome is computed within 1:67 hours. Additional results under various settings can be found in Table 3 and the supplemental material.
15K 相机模型的罗马是计算在 1:67 小时。在不同设置下的其他结果可以在表 3 和补充材料中找到。

The timing in Table 2 shows that BAs (full + partial) account for the majority of the reconstruction time, so the time complexity of the BAs approximates that of the incremental SfM. We keep track of the time of reconstruction and the accumulated time of each step. As shown in Figure 6, the reconstruction time (as well as the accumulated
表 2 中的时间显示，BAs(full+partial)占据了重建时间的大部分，因此 BAs 的时间复杂度接近于嵌入式 SfM。我们记录了重建的时间和每一步的累积时间。如图 6 所示，重建时间(以及积累时间

time of full BAs and partial BAs) increases roughly linearly as the models grow larger, which validates our analysis of the time complexity. The Loop reconstruction has a higher slope mainly due to its higher number of feature matches.
实验结果表明，模型的时间复杂度随着模型的增大而线性增大，验证了我们对模型时间复杂度的分析。循环重建具有较高的斜率，这主要是由于循环特征匹配次数较多。

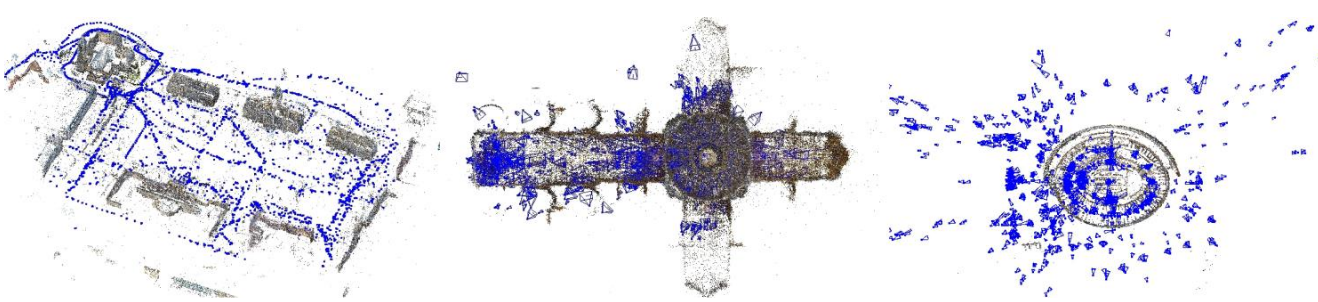
6.3. Reconstruction Quality and Speed
6.3.重建质量和速度

Figure 4, 5 and 7 demonstrate high quality reconstructions that are comparable to other methods, and we have also reconstructed more cameras than DISCO [4] for both
图 4、5 和 7 展示了与其他方法相比的高质量的重构，并且我们还为这两种方法重构了比 DISCO[4]更多的摄像机



(a) Reconstruction time. (b) Accumulated time of full BA. (c) Accumulated time of partial BA.
(a)重建时间。(b)全 BA 的累积时间。(c)部分 BA 的累积时间。

Figure 6. The reconstruction time in seconds as the 3D models grow larger. Here we report the timing right after each full BA. The reconstruction time and the accumulated time of full BAs and partial BAs increase roughly linearly with respect to the number of cameras. Note the higher slope of the Loop reconstruction is due to the higher number of feature matches between nearby video frames.
图 6。随着 3d 模型的增大，重建时间以秒为单位。在这里，我们报告的时间权利后，每个完整的 BA。全场和部分场景的重建时间和累积时间随着摄像机数目的增加大致呈线性增加。注意，Loop 重建的较高斜率是由于邻近视频帧之间特征匹配的次数较多。



(a) Arts Quad (5624 cameras) (b) St. Peter's Basilica (1267 cameras) (c) Colosseum (1157 cameras)
(a)艺术四方(5624 台摄影机)(b)圣彼得大教堂(1267 台摄影机)(c)斗兽场(1157 台摄影机)

Figure 7. Our reconstruction of Arts Quad, St. Peter's Basilica and Colosseum.
图 7.我们对圣彼得大教堂和 Colosseum 艺术广场的重建。

Central Rome and Arts Quad. In particular, Figure 5 shows the correct overlay of our 15065 camera model of Central Rome on an aerial image. The Loop reconstruction shows that the RT steps correctly deal with the drifting problem without explicit loop closing. The reconstruction is first pushed away from local minima by RT and then improved by the subsequent BAs. The robustness and accuracy of our method is due to the strategy of mixed BA and RT.
罗马市中心艺术广场。特别是，图 5 显示了我们的 15065 相机模型的正确覆盖的中央罗马航空图像。循环重构结果表

明，RT 步骤能够正确处理无显式循环关闭的漂移问题。该方法首先利用 RT 推出局部极小点，然后利用后续的 BAs 对重建结果进行改进。该方法的鲁棒性和准确性得益于混合的 BA 和 rt 策略。

We evaluate the accuracy of the reconstructed cameras by comparing their positions to the ground truth locations. For the Arts Quad dataset, our reconstruction (r = 5% and r^0 = 25%) contains 261 out of the 348 images whose ground truth GPS location is provided by [4]. We used RANSAC to estimate a 3D similarity transformation between the 261 camera locations and their Euclidean coordinates. With the best found transformation, our 3D model has a mean error of 2:5 meter and a median error of 0:89 meter, which is smaller than the 1:16 meter error reported by [4].

我们通过比较重建相机的位置和地面真实位置来评估它们的准确性。对于 ArtsQuad 数据集，我们的重建(r5%和 r025%)包含 348 张图像中的 261 张，其地面真实 GPS 定位由[4]提供。我们使用 RANSAC 来估计 261 个相机位置和它们的欧几里德坐标之间的 3d 相似性变换。通过最佳发现转换，我们的 3D 模型的平均误差为 2:5 米，中位误差为 0:89 米，这比文献[4]报告的 1:16 米误差要小。

We evaluate the reconstruction speed by two measures:
我们用两种方法来评估重建速度:

t=n The time needed to reconstruct a camera.
重建照相机所需的时间。

t=q The time needed to reconstruct an observation. Table 3 shows the comparison between our reconstruction under various settings and the DISCO and bundler reconstruction of [4]. While producing comparably large models, our algorithm normally requires less than half a second to reconstruct a camera in terms of overall speed. By

重建一个观察所需的时间。表 3 显示了我们在各种情况下的重建与 DISCO 和 bundler 的重建之间的比较[4]。虽然生成相当大的模型，我们的算法通常需要不到半秒的时间来重建一个摄像机的总体速度。作者

t=n, our method is 8-19X faster than DISCO and 55-163X faster than bundler. Similarly by t=q, our method is 5-11X faster than DISCO and 56-186X faster than bundler. It is worth pointing out that our system uses only a single PC (12GB RAM) while DISCO uses a 200 core cluster.
我们的方法比 DISCO 快 8-19 倍，比 bundler 快 55-163 倍。类似地，我们的方法比 DISCO 快 5-11 倍，比 bundler 快 56-186 倍。值得指出的是，我们的系统只使用一台 PC(12gb 内存)，而 DISCO 使用一个 200 核的集群。

6.4. Discussions
6.4.讨论

Although our experiments show approximately linear running times, the reconstruction takes $O(n^2)$ in theory and will exhibit such a trend the problem gets even larger. In addition, it is possible that the proposed strategy will fail for extremely larger reconstruction problems due to larger accumulated errors. Thanks to the stability of SfM, mixed BAs and RT, our algorithm works without quality problems even for 15K cameras.
虽然我们的实验显示了近似线性的运行时间，但是重构在理论上需要 o(n2)，并且将呈现这样的趋势，问题变得更大。此外，由于较大的累积误差，对于极大的重建问题，提出的策略可能会失败。由于 SfM，混合 BAs 和 RT 的稳定性，我们的算法工作没有质量问题，甚至 15K 相机。

This paper focuses on the incremental SfM stage, while the bottleneck of 3D reconstruction sometimes is the image matching. In order to test our incremental SfM algorithm on the largest possible models, we have matched up to $O(n^2)$ image pairs for several datasets, which is higher than the vocabulary strategy that can choose O(n) pairs to match [2]. However, this does not limit our method from working with fewer image matches. From a different angle, we have contributed the preemptive matching that can significantly reduce the matching cost.
本文重点研究了增量式 SfM 阶段，而三维重建的瓶颈有时是图像匹配。为了在最大可能的模型上测试我们的增量 SfM 算法，我们为多个数据集匹配了最多的 o(n2)图像对，这比可以选择 o(n)对匹配的词汇策略[2]要高。但是，这并不限制我们的方法使用更少的图像匹配。从另一个角度出发，我们提出了抢占式匹配，可以显著降低匹配成本。

Dataset 数据集	Full BA 全文英 文字母	Partial BA 部分 BA	RT 反应时	n	q	t	t=n 不要	t=q 我不知 道你在 说什么
Central Rome 罗马中部	r = 5% R 5%	Every Image 每一张图片	r 0 R = 25% 0 25%	15065	12903K 12903K	1.67hour 1.67 小时	0.40s 0.40 s	0.47ms 0.47 m / s
	r = 25% R 25%	Every Image 每一张图片	r 0 R = 50% 0 50%	15113	12958K 12958K	1.32hour 1.32 小时	0.31s 0.31 s	0.37ms 0.37 m / s
	r = 5% R 5%	Every 3 Images 每 3 张图片	r 0 R = 25% 0 25%	14998	12599K 12599K	1.03hour 1.03 小时	0.25s 0.25 s	0.29ms 0.29 m / s
	DISCO result of [4] 的 DISCO 结果[4]			14754	21544K 21544K	13.2hour 13.2 小时	3.2s 3.2 s	2.2ms 2.2 m / s
	Bundler result of [4] 捆绑器的结果[4]			13455	5411K 5411K	82hour 82 小时	22s 22s	54ms 54ms
Arts Quad 艺术方阵	r = 5% R 5%	Every Image 每一张图片	r 0 R = 25% 0 25%	5624	5839K 5839K	0.59hour 0.59 小时	0.38s 0.38 s	0.37ms 0.37 m / s
	r = 25% R 25%	Every Image 每一张图片	r 0 R = 50% 0 50%	5598	5850K 5850K	0.42hour 0.42 小时	0.27s 0.27 s	0.26ms 0.26 m / s
	r = 5% R 5%	Every 3 Images 每 3 张图片	r 0 R = 25% 0 25%	5461	5530K 5530K	0.53hour 0.53 小时	0.35s 0.35 s	0.35ms 0.35 m / s
	DISCO result of [4] 的 DISCO 结果[4]			5233	9387K 9387K	7.7hour 7.7 小时	5.2s 5.2 s	2.9ms 2.9 m / s
	Bundler result of [4] 捆绑器的结果[4]			5028	10521K 10521K	62hour 62 小时	44s 44 口 径	21ms 21ms
Loop 环路	r = 5% R 5%	Every Image 每一张图片	r 0 R = 25% 0 25%	4342	7196K 7196K	3251s 3251	0.75s 0.75 s	0.45ms 0.45 m / s
	r = 25% R 25%	Every Image 每一张图片	r 0 R = 50% 0 50%	4342	7574K 7574K	1985s 1985 年 的	0.46s 0.46 口径	0.26ms 0.26 m / s
	r = 5% R 5%	Every 3 Images 每 3 张图片	r 0 R = 25% 0 25%	4341	7696K 7696K	3207s 3207	0.74s 0.74 s	0.41ms 0.41 m / s
St. Peter’s 圣彼得教堂	r = 5% R 5%	Every Image 每一张图片	r 0 R = 25% 0 25%	1267	2706K 2706K	583s 583s	0.46s 0.46 口径	0.22ms 0.22 m / s
	r = 25% R 25%	Every Image 每一张图片	r 0 R = 50% 0 50%	1267	2760K 276 万	453s 453s	0.36s 0.36 s	0.16ms 0.16 m / s
	r = 5% R 5%	Every 3 Images 每 3 张图片	r 0 = 25% R 25%	1262	2668K 2668K	367s 367	0.29s 0.29 s	0.14ms 0.14 m / s

			0					
Colosseum 圆形竞技场	r = 5% R 5%	Every Image 每一张图片	r 0 R = 25% 0 25%	1157	1759K 1759K	591s 591	0.51s 0.51 s	0.34ms 0.34 m / s
	r = 25% R 25%	Every Image 每一张图片	r 0 R = 50% 0 50%	1087	1709K 1709K	205s 205 口径 的	0.19s 0.19 s	0.12ms 0.12 m / s
	r = 5% R 5%	Every 3 Images 每 3 张图片	r 0 R = 25% 0 25%	1091	1675K 1675K	471s 471	0.43s 0.43 s	0.28ms 0.28 m / s

Table 3. The statistics of our reconstructions under various settings. We reconstruct larger models than DISCO and bundler under the various settings, and our method also runs significantly faster. Additional results will be presented in the supplemental material.
表三。我们在各种情况下重建的统计数据。在不同的设置下，我们重建了比 DISCO 和 bundler 更大的模型，而且我们的方法也运行得更快。其他的结果将在补充材料中提出。

7. Conclusions and Future Work
7.结论和未来工作

This paper revisits and improves the classic incremental SfM algorithm. We propose a preemptive matching method that can significantly reduce the feature matching cost for large scale SfM. Through algorithmic analysis and extensive experimental validation, we show that incremental SfM is of $O(n^2)$ time complexity, but requires only O(n) time on its major steps while still maintaining the recon-struction accuracy using mixed BAs and RT. The practical run time is approximately O(n) for large problems up to 15K cameras. Our system demonstrates state of the art performance by an average speed of reconstructing about 2 cameras and more than 2000 features points in a second for very large photo collections.

本文对经典的增量 SfM 算法进行了修正和改进。我们提出了一种抢占式匹配方法，可以显著降低大规模 SfM 的特征匹配成本。通过算法分析和大量的实验验证，我们发现增量式 SfM 具有 o(n2)时间复杂度，但在主要步骤上只需要 o(n)时间，而混合式 BAs 和 rt 仍能保持侦察精度。对于 15K 摄像机以下的大型问题，实际运行时间大约为 o(n)。我们的系统展示了最先进的性能，通过平均速度重建约 2 个相机和超过 2000 个功能点在一秒钟为非常大的照片集。

In the future, we wish to explore a more adaptive pre-emptive feature matching and guide the full BAs according to the accumulation of reprojection errors.

今后，我们希望探索一种更具自适应性的先发制人特征匹配方法，并根据重投影误差的累积对全局特征进行引导。

References
参考资料

[1] S. Agarwal, N. Snavely, S. Seitz, and R. Szeliski. Bundle adjustment in the large. In ECCV, pages II: 29–42, 2010. 2, 3

S.Agarwal， n.Snavely， s.Seitz， r.Szeliski. 大型光束法平差。在 ECCV，第二页:29-42,2010。2,3

[2] S. Agarwal, N. Snavely, I. Simon, S. M. Seitz, and R. Szeliski. Building Rome in a day. In ICCV, 2009. 1, 2, 7

西蒙，s.m.Seitz，r.Szeliski。一天建成罗马。在 ICCV，2009 年。1,2,7

[3] M. Byrod and K. Astrom. Conjugate gradient bundle adjustment. In ECCV, pages II: 114–127, 2010. 2, 3

和 k.Astrom.共轭梯度光束法平差.在 ECCV 中，第二页:114-127,2010.2,3

[4] D. Crandall, A. Owens, N. Snavely, and D. P. Huttenlocher. Discrete-continuous optimization for large-scale structure from motion. In CVPR, 2011. 1, 2, 5, 6, 7, 8

克兰德尔，a.欧文斯，n.Snavely，和 d.p.哈滕洛克。大型结构的离散-连续运动优化。在 CVPR，2011 年。1,2,5,6,7,8

[5] J. Frahm, P. Fite Georgel, D. Gallup, T. Johnson, R. Raguram, C. Wu, Y. Jen, E. Dunn, B. Clipp, S. Lazebnik, and
弗拉姆, p.FiteGeorgel, d.Gallup, t.Johnson, r.Ragu-ram, c.Wu, y.Jen, e.Dunn, b.Clipp, s.Lazebnik, and
M. Pollefeys. Building rome on a cloudless day. In ECCV, pages IV: 368–381, 2010. 1, 2
《在晴朗无云的日子里建立罗马》, 载于 ECCV, 第四页:368-381,2010.1,2

[6] R. Gherardi, M. Farenzena, and A. Fusiello. Improving the efficiency of hierarchical structure-and-motion. In CVPR, pages 1594–1600, 2010. 2
和 a.Fusiello。提高层次结构和运动的效率。在 CVPR, 第 1594-1600 页, 2010。图 2

[7] A. Kushal, B. Self, Y. Furukawa, C. Hernandez, D. Gallup,
库沙尔, b.Self, y.Furukawa, c.Hernandez, d.Gallup,
B. Curless, and S. Seitz. Photo tours. In 3DimPVT, 2012. 1
2012.1

[8] X. Li, C. Wu, C. Zach, S. Lazebnik, and J. Frahm. Modeling and recognition of landmark image collections using iconic scene graphs. In ECCV, 2008. 1, 2
李、吴、扎克、拉泽布尼克和弗拉姆。基于标志场景图的地标图像集建模与识别。在 ECCV, 2008 年。1,2

[9] M. A. Lourakis and A. Argyros. SBA: A Software Package for Generic Sparse Bundle Adjustment. ACM Trans. Math. Software, 36(1):1–30, 2009. 3
A.Lourakis 和 a.Argyros。用于通用稀疏光束法平差的软件包。AcmTrans.数学。软件, 36(1):1-30,2009。图 3

[10] D. G. Lowe. Distinctive image features from scale-invariant keypoints. IJCV, 60:91–110, 2004. 2
《尺度不变关键点的独特图像特征》, IJCV, 60:91-110,2004.2

[11] D. Nister and H. Stewenius. Scalable recognition with a vocabulary tree. In CVPR, pages 2161–2168, 2006. 1
D.尼斯特和 H.stevenius。词汇树的可扩展识别。在 CVPR 中, 2161-2168 页, 2006.1

[12] J. R. Shewchuk. An introduction to the conjugate gradient method without the agonizing pain, 1994. 3
没有痛苦的共轭梯度法》, 1994.3

[13] S. N. Sinha, D. Steedly, and R. Szeliski. A multi-stage linear approach to structure from motion. In ECCV RMLE workshop, 2010. 2
斯蒂利和 r.n.辛哈。从运动到结构的多阶段线性方法。在 ECCVRMLE 车间, 2010 年。图 2

[14] N. Snavely, S. Seitz, and R. Szeliski. Photo tourism: exploring photo collections in 3D. In SIGGRAPH, pages 835–846, 2006. 1, 3
和 r.Szeliski。图片旅游:探索 3d 图片收藏。在 SIGGRAPH, 第 835-846 页, 2006 年。1,3

[15] N. Snavely, S. M. Seitz, and R. Szeliski. Skeletal graphs for efficient structure from motion. In CVPR, 2008. 1, 2
史纳夫利, s.m.Seitz, 和 r.Szeliski。从运动中获得有效结构的骨骼图。CVPR, 2008.1,2

[16] C. Wu, S. Agarwal, B. Curless, and S. M. Seitz. Multicore bundle adjustment. In CVPR, 2011. 2, 3
和 s.m.Seitz.多核光束法平差.InCVPR, 2011.2,3