

用于基本矩阵估计的特征匹配算子评价标准

Jia-Wang Bian^{1,2}

jiawang.bian@adelaide.edu.au

Yu-Huan Wu³

wuyuhuan@mail.nankai.edu.cn

Ji Zhao⁴

ji.zhao@tusimple.ai

Yun Liu³

nk12csly@mail.nankai.edu.cn

Le Zhang⁵

zhangleuestc@gmail.com

Ming-Ming Cheng³

cmm@nankai.edu.cn

Ian Reid^{1,2}

ian.reid@adelaide.edu.au

¹School of Computer Science,

The University of Adelaide,

Adelaide, Australia

²Australian Centre for Robotic

Vision, Australia

³Nankai University, China

⁴TuSimple, China

⁵Agency for Science, Technology
and Research, Singapore

摘要

在许多计算机视觉应用中, 估计两个图像的相对几何关系的同时对其进行匹配是关键的一步。数十年来, 由 SIFT、RANSAC 和 8 点算法组成的完整流程常用于此任务。近年来, 许多新方法被提出并在标准测试基准中展示了优于以前方案的性能, 其中包括可学习特征、对应剪枝算法和鲁棒估计。但是, 将以上算法融入传统流程中是否有性能提升却鲜有研究。为此, 我们希望研究 i) 在图像匹配和对极几何估计场景下评价这些最新算法的性能和 ii) 利用这些算法设计更实用的配准系统。我们在四个大规模数据集上使用严格定义的评价标准开展实验, 并且良好的实验结果展示了哪种算法更适合哪种场景。据此, 我们提出了三个高质量匹配系统和 Coarse-to-Fine RANSAC 估计。他们展示了优秀的性能和应用于大量计算机视觉任务的潜力。为了方便将来的研究, 完整的评价流程和我们提出的方法已开源。

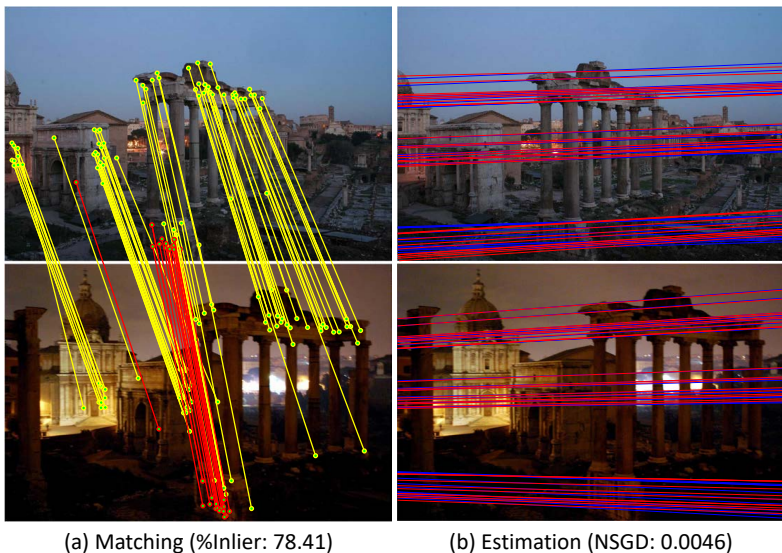


图 1: 两视图匹配和几何估计流程的输出。(a) 为匹配结果, 其中黄色和红色分别表示正确匹配和离群点。(b) 为对极几何估计结果, 其中蓝色和红色分别表示真实数据和估计的极线。我们使用提出的 normalized symmetric geometric distance (NSGD) 来评价估计的准确性。数值越小表示性能越好。

1 引言

匹配两张图片同时恢复他们的几何关系, 如 epipolar geometry [17], 是计算机视觉中最基本的任务之一, 也是许多应用, 如 Structure-from-Motion (SfM) [1, 20, 32, 37, 38] 和 Visual SLAM [11, 15, 31], 的关键步骤。在这些应用中, 整体性能在很大程度上取决于两视图初始配准的质量。因此, 对该模块进行全面的性能评估对于计算机视觉社区至关重要。然而, 据我们所知, 此前没有相关工作。为此, 在本文中, 我们致力于对现有算法进行广泛的实验评估, 以建立统一的评价流程。

数十年来, 一直有一套标准流程应用于此任务, 其中使用 SIFT [24] 特征建立跨图像的初始对应点, 然后使用 Lowe's ratio test [24] 剔除坏的对对应点, 最终使用基于估计算子的 RANSAC [14] 估计几何关系。我们希望恢复基本矩阵 (FM), 该方法比其他几何模型, 如 homography 和本征矩阵 (essential matrix), 更适用于一般化的场景。Fig. 1展示了该流程的输出样例。在此, 我们主要关心几何估计的质量。

最近, 许多新方法被提出, 这些方法显示出在该项任务上的潜力, 包括可学习特征的方法 [25, 29, 30]、鲁棒估计 [6, 34] 和此前受到较少关注的对应点剪枝

算法 [7, 26, 47]。然而, 尽管这些算法的性能在标准测试基准上优于早期算法, 但将它们整合到经典流程中可能未必会带来性能提升。例如, Balntas et al. [4] 展示了在标准测试基准 [8] 上性能优于其他方法的特征描述算子并没有带来更好的匹配质量。在 [7, 39, 47] 中同样展示和讨论了这种不一致性。

在本文中, 我们将最近提出的算法整合到完善的图像匹配和对极几何估计流程中, 以研究它们是否可以提高整体性能, 从而对它们进行全面评估。详细而言, 本文主要做了如下贡献:

- i) 我们提出了一套评估流程用于评价局部特征、鲁棒估计和此前未被仔细研究的对应点剪枝算法如 [7, 26, 47]。
- ii) 我们使用严格定义的指标评估在四个大型数据集上评价算法。结果展示了哪些数据集更有挑战性和哪些算法适用于哪些场景。The results provide insights into which datasets are particularly challenging and which algorithms suit which scenarios.
- iii) 基于结果, 我们提出三个高质量、高效率的匹配系统, 其性能与优秀的 CODE[23] 系统相当但快几个数量级。
- iv) 有趣的是, 我们发现最近的 GC-RANSAC[6] (又称 USAC [34]) 并不能在几何估计中持续取得较高的性能, 但是可以有效地进行离群点裁剪。因此我们提出首先使用该算法进行离群点剔除, 然后应用基于估计算子的 LMedS[36] 进行模型拟合。得到的方法, 又称 Coarse-to-Fine RANSAC, 与其他方法相比具有明显的优越性。

2 相关工作

大量科研工作集中在评价局部特征和鲁棒估计, 而对应点裁剪算法并未得到很好的评估。我们提出的基准测试弥补了这一缺失。接下来我们将介绍具有代表性的评价流程和我们的评价流程。

评估局部特征。 Mikolajczyk 等人 [28] 在小规模数据集上评估了仿射区域检测器, 该数据集涵盖了各种光度和几何图像转换。随后, Mikolajczyk 和 Schmid[27] 将该工作扩展到局部描述算子上。基于此, Heinly 等人 [19] 提出了几个额外的标准和数据集来评估二元描述算子。此外, Brown 等人 [8] 提出了图像块对分类测试基准用于评价可学习描述算子, 该基准测试了描述算子从图像块错误配对中分辨出正确配对的能力。最近, Balntas 等人 [5] 评价了人为设计的和可学习的

Algorithm 1 Compute SGD

Input: I_1, I_2, F_1, F_2, N Output: sgd 1: function ComputeSGD(I_1, I_2, F_1, F_2, N) 2: $sgd \leftarrow 0$ 3: $count \leftarrow 0$ 4: while $count < N$ do 5: randomly choose a point m in I_1 6: draw $l_1 = F_1 m$ in I_2 7: if l_1 does not intersect with I_2 then 8: continue 9: end if 10: randomly choose a point m' on L_1	11: draw $l_2 = F_2 m$ in I_2 12: $d'_1 = \text{distance}(m', l_2)$ 13: draw $l_3 = F_2^T m'$ in I_1 14: $d_1 = \text{distance}(m, l_3)$ 15: $sgd \leftarrow sgd + d'_1 + d_1$ 16: $count \leftarrow count + 1$ 17: end while 18: swap (I_1, I_2), swap (F_1, F_2) 19: repeat step 3 – 17 20: $sgd \leftarrow sgd / (4 * N)$ 21: return sgd 22: end function
--	---

描述算子在验证和检索 homography 块方面的能力。Schönberger 等人 [39] 在基于图像的重建任务中对比评估了这两种类型的描述符。

评估鲁棒估计。 Choi 等人 [10] 评估了 RANSACRANSAC [14] 类方法的线拟合和单应性估计 [17] 能力, 评价指标包括准确率、运行时间和鲁棒性。Lacey 等人 [22] 对用于双目相机校准的 RANSAC 算法进行了评估。Raguram 等人 [33] 对 RANSAC 算法进行了分类, 并对它们进行了比较分析, 其中考虑了效率和准确性之间的权衡。这些评价流程在合成数据和真实数据上评估了鲁棒模型拟合技术。Torr 等人 [42, 44] 提供了基本矩阵 (FM) 估计算法的性能表征。Zhang [48] 回顾了 FM 估计技术并提出了有充分根据的评价标准来计算两个基本矩阵的距离, 该标准被证明比使用 Frobenius 范式更好。Armanguè 等人 [3] 提供了对不同 FM 估计方法的综述。Fathy 等人 [13] 研究了 FM 估计中的误差标准。

提出的基准测试。 我们的基准测试主要受 [39] 启发, 该方法评价了高阶任务中的描述符。不同之处在于, 我们在两视图图像匹配和几何估计任务下评价了三种算法的整体性能, 而 [39] 在多种任务中评价描述符以寻找通用描述符。此外, 我们参考 [6, 35, 39, 42, 48] 来设计评价标准和构建基准测试数据集。而且, 提出的评价标准也可用于对图像匹配和几何估计流程进行消融实验。该标准可以帮助研究人员设计更实用的对应匹配系统。

3 评价标准

3.1 FM 估计评价标准

基本矩阵由于其结构而无法直接比较。为了测量估计的准确性，我们按照 Zhang 的方法 [48]，在本文中被称为对称几何距离 (SGD)。它使用真实的 FM 生成虚拟的对应点，并计算与估计的对应点之间的对极距离，然后对换两者再次计算距离以确保对称性。平均距离被用作准确性度量。Alg. 1 代表计算 SGD 误差的整体流程，其中 (I_1, I_2) 是一对图像对， F_1 和 F_2 是两个基本矩阵， N 是最大迭代次数。

归一化 SGD。 计算得到的 SGD 误差（像素级）会导致具有不同分辨率的图像之间出现可比性问题。为了解决该问题，我们通过将距离除以图像对角线长度，将距离归一化到 $[0, 1]$ 的范围。形式上，距离通过乘以因子 $f = 1/\sqrt{h^2 + w^2}$ 进行正则化，其中 h 和 w 分别代表图像的高度和宽度。这使得该误差可以在不同分辨率的图像之间比较。

% 召回率。 给定 FM 估计，我们通过对归一化 SGD 误差使用不同的阈值将其分类为正确的或错误的，并且使用 % 召回率（正确的估计占全部估计的比例）进行评估。在我们的实验中，0.05 为阈值。由于召回率随阈值的累积而增加，性能对阈值的选择不敏感。不过，我们还是建议读者给出随着阈值变化的召回率曲线。

3.2 图像匹配评价标准

% 正确匹配率。 我们使用正确匹配率（即正确匹配占有所有匹配的比例）来评价匹配质量。在此，在两个图像中距真实对极线的距离小于某个阈值的匹配被视为正确匹配。为了避免不同分辨率图像的可比较性问题，我们将阈值设为 $\alpha\sqrt{h^2 + w^2}$ ，其中 h 和 w 分别为图像的高度和宽度。在我们的评价标准中 α 为 0.003。此外，为了评价中间结果，我们还提出 %Inlier-m，即使用 RANSAC[14] 等鲁棒估计算法进行离群点剔除前的正确匹配率。这反映了纯特征匹配系统的性能。

#Corrs。 我们使用对应点数量而不是性能比较来分析结果，因为匹配数量的影响对 SfM[37] 等高级应用的影响是有争议的。然而，太少的对应点数会使这些应

用退化。因此, 只要匹配数量不是很少我们就不关注匹配数。相似地, #Corrs-m (即估计阶段前的匹配个数) 也被用来分析结果。

4 数据集

我们使用四个大规模基准测试数据集进行评估, 其中包括不同的真实世界场景, 而且相机参数各不相同。这种多样性使我们能够在不同场景下比较算法。

数据集。 基准测试数据集包括: (1) TMU SLAM 数据集 [40], 其中包括室内场景的视频, 由于相机的快速移动, 视频中纹理特征通常较弱并且图像有时是模糊的。(2) KITTI 里程计数据集 [16], 其中包括驾驶情境下连续的帧, 图像间主要的几何关系与向前运动有关。(3) Tanks and Temples(T&T) 数据集 [21], 提供许多场景或物体的扫描, 以用于基于图像的重建, 因此可提供宽基线图像对用于评估。(4) Community Photo Collection (CPC) 数据集 [46], 包括收集自 Flickr 的世界各地著名地标的非结构化图像。在 CPC 数据集中, 图像由不同时间的任意相机拍摄。Fig. 2展示了这些基准测试数据集中的样例图像。

真实数据。 不同图像对间的基本矩阵可以从其投影矩阵 (P 和 P') 代数推导如下:

$$F = [P'C]_{\times} P'P^{+} \quad (1)$$

其中 P^{+} 是 P 的伪逆, 即 $PP^{+} = I$, C 是空向量, 即相机中心, 由公式 $PC = 0$ 定义。 $P = K[R|t]$ 是一个 3×4 矩阵, 并且满足以下公式

$$d \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = P \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \quad (2)$$

其中 d 为未知深度, $[u, v]$ 是图像坐标, $[x, y, z]$ 是真实世界坐标。 K 是相机内参, $[R|t]$ 是相机外参。 TUM 和 KITTI 数据集中有真实的相机内参和相机外参, 而 T&T 和 CPC 数据集中相机内参和外参未知。 因此, 我们按照 [35, 47] 使用 COLMAP[37] 重建图像序列来生成真实相机参数。 注意 SfM 流程对 3D 点和相机的一致性进行全局推理, 因此可以得到平均重投影误差低于一个像素点的准确估计 [37]。

图 2: 基准测试数据集集中的样例图像。

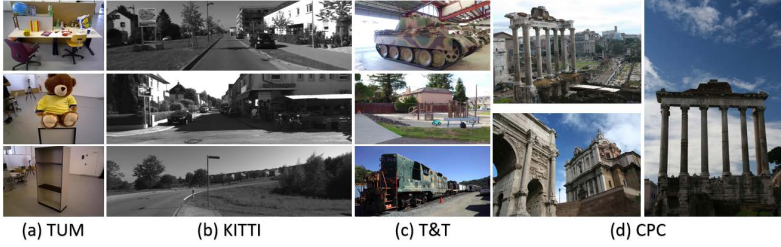


表 1: 基准测试数据集细节。

Datasets	#Seq	#Image	Resolution	Baseline	Property
TUM	3	5994	480×640	short	indoor scenes
KITTI	5	9065	370×1226	short	street views
T&T	3	922	1080×2048 1080×1920	wide	outdoor scenes
CPC	1	1615	varying	wide	internet photos

图像对构建。 我们通过正确匹配数寻找可配对的图像，即我们使用 SIFT[24] 在两张图像间生成对应点并按照 [35] 选择包括至少 20 对正确匹配的图像对。对宽基线数据集 (T&T [21] 和 CPC [46])，在所有图像间搜寻图像对。对短基线数据集 (TUM [40] 和 KITTI [16])，每一帧与接下来一秒内的后续帧配对，因为其他帧间几乎没有重叠。通过这种方式，我们获得了大量的可配对图像对，然后在每个数据集中随机选取 1000 对用作测试。对每个数据集的测试集拆分情况的如下所述。在 TUM 数据集 [40] 中，我们在三个序列上测试方法：fr3/teddy，fr3/large_cabinet 和 fr3/long_office_household。在 KITTI 里程计数据集 [16] 中，使用序列 06-10。在 T&T 数据集 [21] 中，使用序列 Panther，Playground 和 Train。在 CPC 数据集 [46] 中，使用 Roman Forum。其他图像序列可被用作训练数据进一步训练其他基于学习的方法。Tab. 1 总结了我們用作评价的测试集。

5 实验

相关实验非常丰富，所有我们在本文中主要集中在测试最近提出的算法和广泛使用的方法。接下来，我们介绍实验设置，讨论结果并提出我们的方法。

表 3: 待评测算法耗时。

Device	Runtime (seconds)					
L	SIFT	DSP-SIFT	RootSIFT-PCA	L2Net	HardNet++	HesAffNet
	0.702	1.762	0.705	2.260	0.002	0.367
W	LPM	GMS	LC	CODE		
	0.003	0.001	0.021	4.068		
	RANSAC	LMedS	MSAC	USAC	GC-RSC	
	0.521	0.528	0.537	0.565	0.788	

5.1 实验设置

基线和对比如。 我们使用经典流程作为基线模型。具体地说, 我们使用 DoG[24] 检测器和 SIFT[24] 描述符通过最朴素的最近邻搜索生成跨图像的初始匹配, 然后使用 Lowe 的 ratio test[24] 裁剪错误匹配, 最终使用 RANSAC[14] 和 8 点算法 [18] 估计 FM 并剔除离群点。对每个被评估的算法, 我们将其整合进基线系统中来替换其对应部分, 然后使用整体性能进行比较。

待评价方法。 首先, 我们评价四个基于学习的局部特征, 包括 HesAffNet[30] 检测算子和两个描述符 (L2Net [41], HardNet++ [29])。此外, 我们也评价了两个人工设计的描述符 (DSP-SIFT [12] 和 RootSIFT-PCA [2, 9]), 其在最近的基准测试 [39] 中展现了较高的性能。其次, 我们评估四个对应点裁剪算法: CODE[23], GMS[7], LPM[26] 和 LC[47]。最后, 我们评估了两个广泛使用的估计算法 (LMedS [36] 和 MSAC [43]) 和两个最新的算法 (USAC [34] 和 GC-RANSAC [6])。

实现。 我们使用 VLFeat [45] 库实现 SIFT 描述符和 DoG 检测算子, ratio test [24] 的阈值设为 0.8。使用 Matlab 函数实现 RANSAC, LMedS 和 MSAC, 其中我们为了速度考虑限制最大迭代轮数为 2000。其他代码取自相应作者的开源实现, 其中基于学习的方法我们使用作者开源的预训练模型。

5.2 结果与讨论

Tab. 2 展示了所有方法和数据集的实验结果。在每个模块中, 第一行展示了基线模型性能。**第一, 第二, 第三**结果使用不同颜色标注, 优于基线性能的结果使用粗体标注。在此, 我们主要比较算法的 %Recall, 其反映了整体性能。此外, %Inlier 表示匹配性能, %Inlier-m 表示剔除离群点前的匹配性能。#Corrs(-m) 用来分析结果而不是性能比较。关于这些度量标准的细节见于 Sec. 3。对于性能

分析, 我们的主要目标是总结最优方法的特性, 而不是对所有方法进行全面比较。

局部特征。 Tab. 2(a) 展示了局部特征的结果。%Recall 显示 a) RootSIFT-PCA [9] 和 HardNet++ [29] 性能持续超过基线模型, 并且后者优于前者。b) HesAffNet [30] 在宽基线场景 (T&T 和 CPC) 中性能最优, 尽管它在 TMU 数据集上结果会退化。c) DSP-SIFT [12] 在除了 TUM 外所有的数据集上性能超过基线模型, L2Net [41] 在所有数据集上与基线模型性能相当。

对应点裁剪方法。 Tab. 2(b) 展示了裁剪方法的结果。由表可见 a) CODE [23] 在所有数据集上取得最优性能。b) GMS [7] 和 LPM [26] 由于高效地裁剪错误匹配点性能持续超过基线模型, 既改善 %Inlier-m 同时保留适当的 #Corrs-m。因此, GMS 优于 LPM。c) LC [47] 可以提高匹配准确率 (%Inlier-m), 但是对于估计 (%Recall) 会在短基线数据集 (TUM 和 KITTI) 上退化。这可能是由于提供的模型是在宽基线数据集上训练得到的。同时, 注意该模型需要相机内参, 相机内参对于 FM 估计问题通常假设是未知的。

鲁棒估计。 Tab. 2(c) 展示了鲁棒估计的结果。由表可见: a) LMedS [36] 在前三个数据集上性能最优, 其中图片比 CPC 数据集更简单。这证实了 Matlab 文档的建议 LMedS 在正确匹配率足够高时 (如高于 50%) 效果更好。b) GC-RANSAC [6] 和 USAC [34] 在宽基线场景下效果很好, 尤其在具有挑战性的 CPC 数据集上。然而, 该方法在短基线场景 (TUM 和 KITTI) 下性能退化。c) 有趣的是, 我们观察到 GC-RANSAC (又称 USAC) 能有效地保留丰富的对应点 (%Corrs-m) 并剔除离群点 (%Inlier(-m))。

运行时间 由于算法依赖不同的操作系统, 我们使用两台机器进行评估: 一台 Linux 服务器 L (Intel E5-2620 CPU, NVIDIA Titan Xp GPU) 和 Windows 笔记本 W (Intel i7-3630QM CPU, NVIDIA GeForce GT 650M GPU), 其中使用 100 张 KITTI 数据集中的图片进行测试并展示平均结果。Tab. 3 展示了算法的运行时间。描述符依赖 DoG [24] 检测器和 HesAffNet [30], 以上检测器耗时 238ms 分别可提取 1760 关键点和 4860 关键点。CODE [23] (W) 使用 GPU 耗时 2.953s 提取 58,675 关键点, 并在 CPU 上耗时 4.068s 裁剪错误匹配。

5.3 提出的方法

受结果启发, 我们提出如下三个可行的匹配系统和一个鲁棒估计算法。

表 4: 所提出的匹配系统的评估结果。

Datasets	Methods	%Recall	%Inlier	%Inlier-m	#Corrs(-m)
TUM	Baseline	57.40	75.33	59.21	65 (316)
	CODE	67.50	76.04	66.82	9281 (18562)
	RootSIFT-PCA + GMS	67.50	76.13	69.62	124 (248)
	HardNet + GMS	68.60	75.85	69.39	128 (256)
	HesAffNet + GMS	66.40	75.92	67.04	288 (577)
KITTI	Baseline	91.70	98.20	87.40	154 (525)
	CODE	91.90	98.22	93.03	9623 (19246)
	RootSIFT-PCA + GMS	92.50	98.54	95.73	225 (450)
	HardNet + GMS	92.10	98.49	95.43	236 (472)
	HesAffNet + GMS	91.80	98.48	94.18	540 (1079)
T&T	Baseline	70.00	75.20	53.25	85 (795)
	CODE	92.70	87.81	76.98	4626 (9251)
	RootSIFT-PCA + GMS	89.30	85.29	78.69	307 (614)
	HardNet + GMS	92.20	85.52	78.86	343 (686)
	HesAffNet + GMS	90.90	86.16	79.25	412 (824)
CPC	Baseline	29.20	67.14	48.07	60 (415)
	CODE	61.80	89.45	78.55	2890 (5774)
	RootSIFT-PCA + GMS	57.30	88.94	83.70	133 (263)
	HardNet + GMS	60.10	88.34	83.12	149 (298)
	HesAffNet + GMS	60.80	88.72	83.16	182 (362)

匹配系统。 我们首先采用以下三对检测器和描述符中的一对生成参考对应点:

1. DoG [24] + RootSIFT-PCA [9]
2. DoG + (HardNet++) [29]
3. HesAffNet [30] + (HardNet++)

我们推荐 1, 2 用于一般场景 3 用于宽基线场景。然后, 我们应用 ratio test (阈值设为 0.8) 和 GMS [7] 裁剪错误匹配。最后, 我们使用基于 LMedS [36] 的估计算法以拟合模型。Tab. 4展示了评估结果, 表格清晰地展示了推荐的系统性能超过了基线模型, 并且得到了与最先进系统 (CODE [23] + LMedS [36]) 相当的性能。注意 CODE 慢几个数量级, 即使使用 GPU。

Coarse-to-Fine RANSAC. 虽然 Tab. 2展示了 GC-RANSAC [6] 和 USAC [34] 有效地剔除离群点, 但是它们在模型拟合时不能持续取得较高的性能。此外, 我们提出使用 GC-RANSAC [6] 进行错误匹配剔除, 然后使用基于 LMedS [36] 的

表 5: 所提出的 CF-RSC 的 %Recall

Datasets	RANSAC	LMedS	MSAC	USAC	GC-RSC	CF-RSC
TUM	57.40	69.20	52.70	56.50	30.80	69.30
KITTI	91.70	91.80	91.80	82.70	56.50	92.30
T&T	70.00	83.40	64.60	78.80	80.40	90.70
CPC	29.20	44.00	23.00	49.70	53.70	60.90

估计算子进行模型拟合。注意 USAC 也适用。在这个两阶段框架中，前者用来粗略地寻找正确匹配集，后者用来准确地拟合模型，所以我们命名以上算法为 Coarse-to-Fine RANSAC (简称 CF-RSC)。Tab. 5 使用 %Recall 展示了所提方法的结果，其中所有的估计算子使用相同的输入，即 SIFT [24] 也适用于 ratio test 剪枝。该表说明所提出的 CF-RSC 大幅度超过其他方法。

6 结论

该文使用严格定义的标准，在图像匹配和基础矩阵估计任务上，评估了最近提出的局部特征、对应点剪枝算法和鲁棒估计算法。在四个大规模数据集上的综合评价结果显示了哪个数据集极具挑战性和哪种算法更适用于何种场景。这可以促进相关领域研究的进步，并且可以帮助研究人员在不同应用中设计更实用的匹配系统。最后，受结果启发，我们提出三个高质量图像匹配系统和鲁棒估计算法 Coarse-to-Fine RANSAC。以上算法取得了出色的性能并且有广泛应用于计算机视觉任务的潜力。

7 致谢

作者希望感谢图森和华为技术有限公司。

References

- [1] Sameer Agarwal, Yasutaka Furukawa, Noah Snavely, Ian Simon, Brian Curless, Steven M Seitz, and Richard Szeliski. Building Rome in a day. *Communications of the ACM*, 54(10):105–112, 2011.
- [2] Relja Arandjelovic and Andrew Zisserman. Three things everyone should

-
- know to improve object retrieval. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 2911–2918. IEEE, 2012.
- [3] Xavier Armangué and Joaquim Salvi. Overall view regarding fundamental matrix estimation. *Image and Vision Computing*, 21(2):205–220, 2003.
- [4] Vassileios Balntas, Edgar Riba, Daniel Ponsa, and Krystian Mikolajczyk. Learning local feature descriptors with triplets and shallow convolutional neural networks. In *British Machine Vision Conference (BMVC)*, page 3, 2016.
- [5] Vassileios Balntas, Karel Lenc, Andrea Vedaldi, and Krystian Mikolajczyk. HPatches: A benchmark and evaluation of handcrafted and learned local descriptors. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5173–5182. IEEE, 2017.
- [6] Daniel Barath and Jiri Matas. Graph-Cut RANSAC. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.
- [7] JiaWang Bian, Wen-Yan Lin, Yasuyuki Matsushita, Sai-Kit Yeung, Tan Dat Nguyen, and Ming-Ming Cheng. GMS: Grid-based motion statistics for fast, ultra-robust feature correspondence. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4181–4190. IEEE, 2017.
- [8] Matthew Brown, Gang Hua, and Simon Winder. Discriminative learning of local image descriptors. *IEEE Transactions on Pattern Recognition and Machine Intelligence (PAMI)*, pages 43–57, 2011.
- [9] Andrei Bursuc, Giorgos Toliás, and Hervé Jégou. Kernel local descriptors with implicit rotation matching. In *International Conference on Multimedia Retrieval*, pages 595–598. ACM, 2015.
- [10] Sunglok Choi, Taemin Kim, and Wonpil Yu. Performance evaluation of RANSAC family. *International Journal on Computer Vision (IJCV)*, 24(3): 271–300, 1997.
- [11] Andrew J Davison, Ian D Reid, Nicholas D Molton, and Olivier Stasse. MonoSLAM: Real-time single camera slam. *IEEE Transactions on Pattern Recognition and Machine Intelligence (PAMI)*, 29(6):1052–1067, 2007.

-
- [12] Jingming Dong and Stefano Soatto. Domain-size pooling in local descriptors: DSP-SIFT. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 5097–5106, 2015.
- [13] Mohammed E Fathy, Ashraf S Hussein, and Mohammed F Tolba. Fundamental matrix estimation: A study of error criteria. *Pattern Recognition Letters*, 32(2):383–391, 2011.
- [14] Martin A Fischler and Robert C Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.
- [15] Christian Forster, Matia Pizzoli, and Davide Scaramuzza. SVO: Fast semi-direct monocular visual odometry. In International Conference on Robotics and Automation (ICRA), pages 15–22. IEEE, 2014.
- [16] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the KITTI vision benchmark suite. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 3354–3361. IEEE, 2012.
- [17] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.
- [18] Richard I Hartley. In defense of the eight-point algorithm. *IEEE Transactions on Pattern Recognition and Machine Intelligence (PAMI)*, 19(6):580–593, 1997.
- [19] Jared Heinly, Enrique Dunn, and Jan-Michael Frahm. Comparative evaluation of binary features. In European Conference on Computer Vision (ECCV), pages 759–773. Springer, 2012.
- [20] Jared Heinly, Johannes L Schonberger, Enrique Dunn, and Jan-Michael Frahm. Reconstructing the world* in six days*(as captured by the yahoo 100 million image dataset). In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 3287–3295, 2015.
- [21] Arno Knapitsch, Jaesik Park, Qian-Yi Zhou, and Vladlen Koltun. Tanks and Temples: Benchmarking large-scale scene reconstruction. *ACM Transactions on Graphics (TOG)*, 36(4):78, 2017.

-
- [22] AJ Lacey, N Pinitkarn, and Neil A Thacker. An evaluation of the performance of RANSAC algorithms for stereo camera calibration. In *British Machine Vision Conference (BMVC)*, pages 1–10, 2000.
- [23] Wen-Yan Lin, Fan Wang, Ming-Ming Cheng, Sai-Kit Yeung, Philip HS Torr, Minh N Do, and Jiangbo Lu. CODE: Coherence based decision boundaries for feature correspondence. *IEEE Transactions on Pattern Recognition and Machine Intelligence (PAMI)*, 2017.
- [24] David G Lowe. Distinctive image features from scale-invariant keypoints. *International Journal on Computer Vision (IJCV)*, 60(2):91–110, 2004.
- [25] Zixin Luo, Tianwei Shen, Lei Zhou, Siyu Zhu, Runze Zhang, Yao Yao, Tian Fang, and Long Quan. GeoDesc: Learning local descriptors by integrating geometry constraints. In *European Conference on Computer Vision (ECCV)*, 2018.
- [26] Jiayi Ma, Ji Zhao, Junjun Jiang, Huabing Zhou, and Xiaojie Guo. Locality preserving matching. *International Journal on Computer Vision (IJCV)*, 2018.
- [27] Krystian Mikolajczyk and Cordelia Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Recognition and Machine Intelligence (PAMI)*, 27(10):1615–1630, 2005.
- [28] Krystian Mikolajczyk, Tinne Tuytelaars, Cordelia Schmid, Andrew Zisserman, Jiri Matas, Frederik Schaffalitzky, Timor Kadir, and Luc Van Gool. A comparison of affine region detectors. *International Journal on Computer Vision (IJCV)*, 65(1-2):43–72, 2005.
- [29] Anastasiia Mishchuk, Dmytro Mishkin, Filip Radenovic, and Jiri Matas. Working hard to know your neighbor’s margins: Local descriptor learning loss. In *Neural Information Processing Systems (NIPS)*, pages 4826–4837, 2017.
- [30] Dmytro Mishkin, Filip Radenovic, and Jiri Matas. Repeatability is not enough: Learning affine regions via discriminability. In *European Conference on Computer Vision (ECCV)*. Springer, 2018.

-
- [31] Raul Mur-Artal, Jose Maria Martinez Montiel, and Juan D Tardos. ORB-SLAM: a versatile and accurate monocular slam system. *IEEE Transactions on Robotics (TRO)*, 31(5):1147–1163, 2015.
- [32] Filip Radenovic, Johannes L Schönberger, Dinghuang Ji, Jan-Michael Frahm, Ondrej Chum, and Jiri Matas. From dusk till dawn: Modeling in the dark. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5488–5496, 2016.
- [33] Rahul Raguram, Jan-Michael Frahm, and Marc Pollefeys. A comparative analysis of RANSAC techniques leading to adaptive real-time random sample consensus. In *European Conference on Computer Vision (ECCV)*, pages 500–513. Springer, 2008.
- [34] Rahul Raguram, Ondrej Chum, Marc Pollefeys, Jiri Matas, and Jan-Michael Frahm. USAC: a universal framework for random sample consensus. *IEEE Transactions on Pattern Recognition and Machine Intelligence (PAMI)*, 35(8):2022–2038, 2013.
- [35] René Ranftl and Vladlen Koltun. Deep fundamental matrix estimation. In *European Conference on Computer Vision (ECCV)*, pages 284–299, 2018.
- [36] Peter J Rousseeuw and Annick M Leroy. *Robust regression and outlier detection*, volume 589. John wiley & sons, 1987.
- [37] Johannes L Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4104–4113, 2016.
- [38] Johannes L Schönberger, Filip Radenovic, Ondrej Chum, and Jan-Michael Frahm. From single image query to detailed 3d reconstruction. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5126–5134, 2015.
- [39] Johannes L Schönberger, Hans Hardmeier, Torsten Sattler, and Marc Pollefeys. Comparative evaluation of hand-crafted and learned local features. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6959–6968. IEEE, 2017.

-
- [40] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers. A benchmark for the evaluation of RGB-D SLAM systems. In *IEEE International Conference on Intelligent Robots and Systems (IROS)*, Oct. 2012.
- [41] Yurun Tian, Bin Fan, Fuchao Wu, et al. L2-Net: Deep learning of discriminative patch descriptor in euclidean space. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, page 6, 2017.
- [42] Philip HS Torr and David W Murray. The development and comparison of robust methods for estimating the fundamental matrix. *International Journal on Computer Vision (IJCV)*, 24(3):271–300, 1997.
- [43] Philip HS Torr and Andrew Zisserman. MLESAC: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding (CVIU)*, 78(1):138–156, 2000.
- [44] Philip HS Torr and A Zissermann. Performance characterization of fundamental matrix estimation under image degradation. *Machine Vision and Applications*, 9(5-6):321–333, 1997.
- [45] Andrea Vedaldi and Brian Fulkerson. VLFeat: An open and portable library of computer vision algorithms. In *ACM International Conference on Multimedia (ACM MM)*, pages 1469–1472. ACM, 2010.
- [46] Kyle Wilson and Noah Snavely. Robust global translations with 1DSFM. In *European Conference on Computer Vision (ECCV)*, pages 61–75. Springer, 2014.
- [47] Kwang Moo Yi, Eduard Trulls, Yuki Ono, Vincent Lepetit, Mathieu Salzmann, and Pascal Fua. Learning to find good correspondences. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.
- [48] Zhengyou Zhang. Determining the epipolar geometry and its uncertainty: A review. *International Journal on Computer Vision (IJCV)*, 27(2):161–195, Mar 1998. ISSN 1573-1405. doi: 10.1023/A:1007941100561.