



三维视觉研究及应用

吴毅红

2017年5月22日



提 纲

- 概念介绍
- 应用
- 图像匹配：
 - L2-Net: 深度学习图像描述子
- 视觉定位：
 - SLAM重定位技术
 - OpticalFlowSLAM
- 三维重建：
 - HSfM: Hybrid Structure from Motion
 - 大场景三维重建
- 展望与趋势

概念介绍

计算机视觉

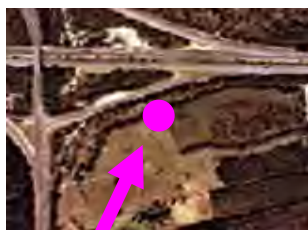
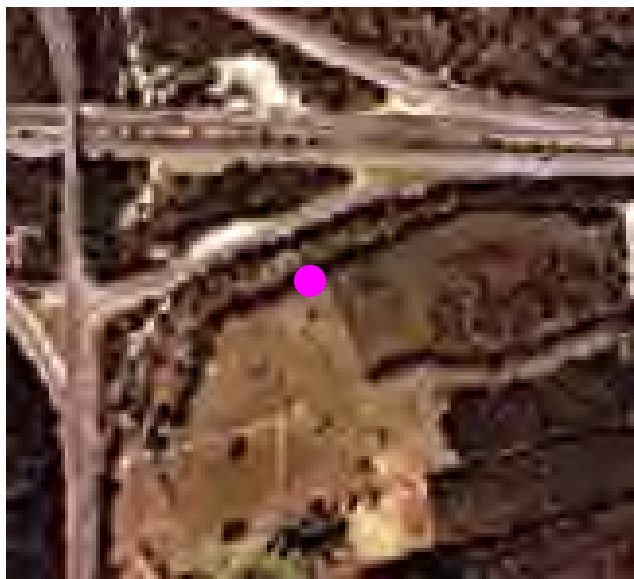
- 用计算机模拟人或生物视觉功能的技术学科
- “眼睛”：相机
- 输入：图像或视频



- 目标：使计算机具有通过图像或视频认知周围环境信息的能力
- 三维计算机视觉：感知周围的三维环境、物体的形状大小位置姿态。
- 三个关键的任务：

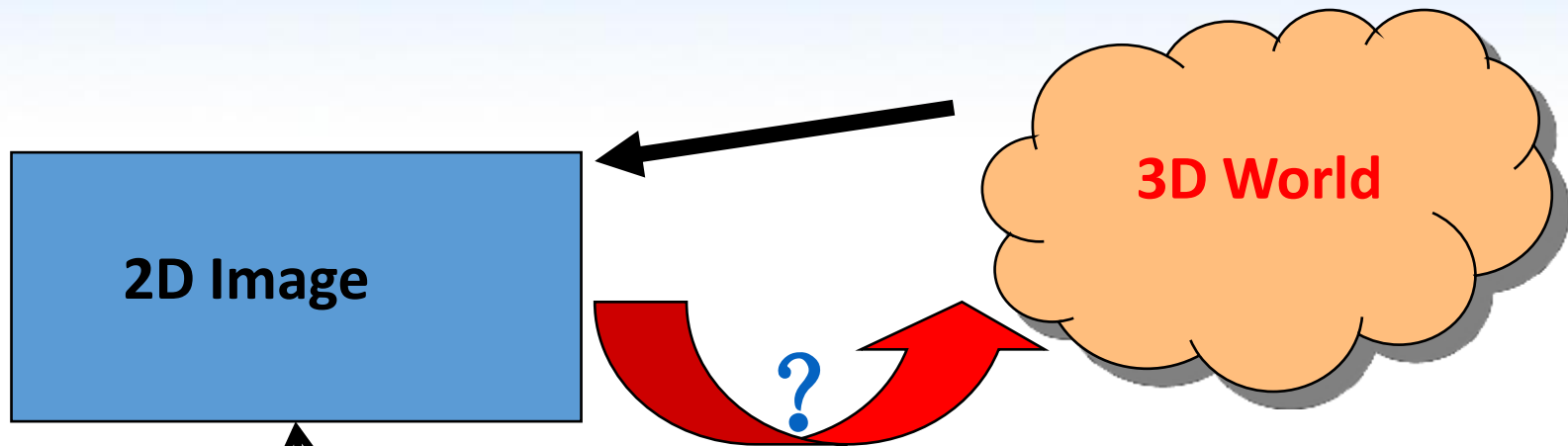
图像匹配、相机定位、三维重建

什么是图像匹配?



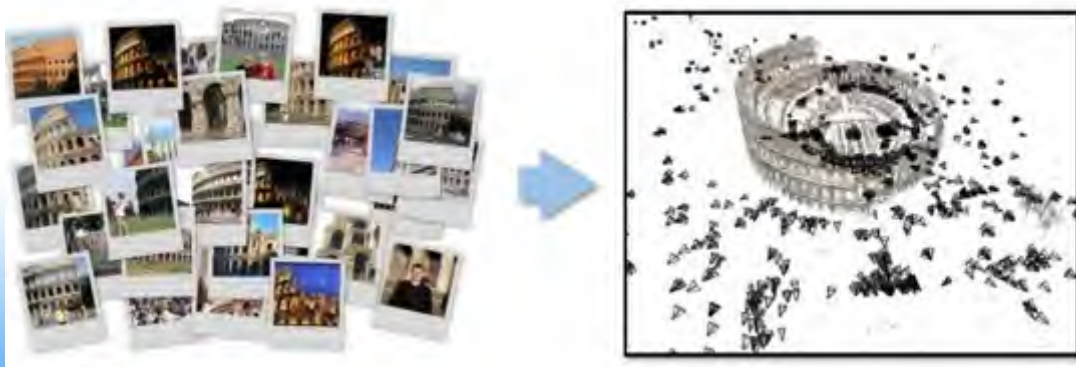
对应点

什么是相机定位?



相机拍摄的2D图像或视频

根据输入的图像或视频，计算出相机在三维物理空间中的位置和姿态称为相机定位



图片来自:

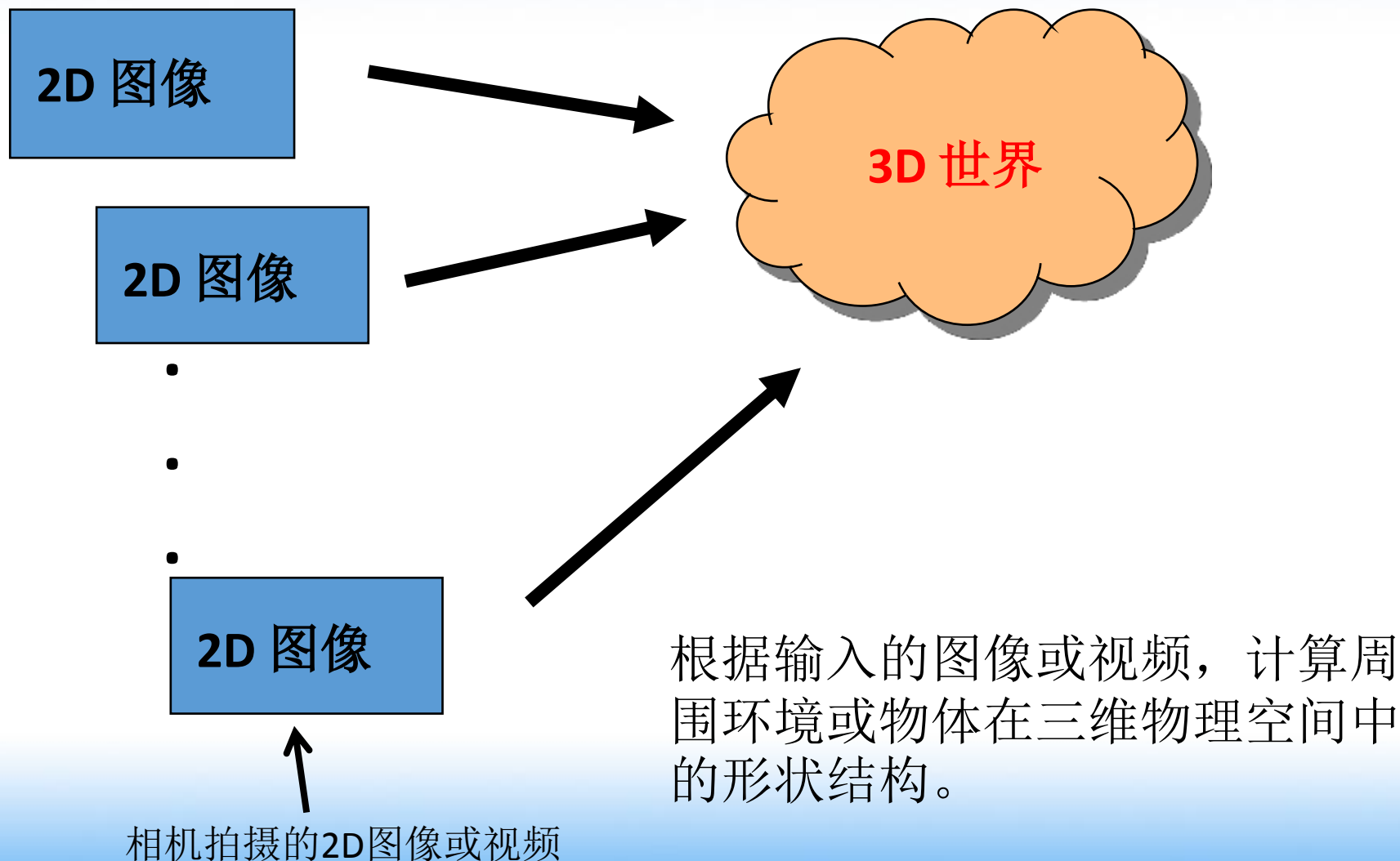
Noah Snavely, Steven M. Seitz, Richard Szeliski.
Modeling the World from Internet Photo
Collections. IJCV 2007.

Sameer Agarwal, Yasutaka Furukawa, Noah
Snavely, Brian Curless, Steven M. Seitz, Richard
Szeliski. Reconstructing Rome. *Computer*, 2010.

关键问题:

- 匹配对应
- 相机标定
- 位姿计算

什么是基于图像的三维重建?



关键问题

- 相机标定
- 图像匹配
- 相机定位
- 三角化
- 优化

应用

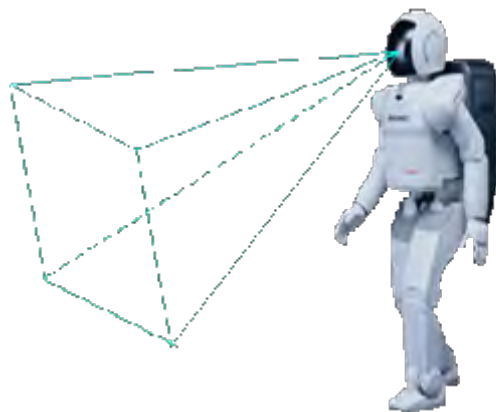
- 机器人导航，无人驾驶，辅助驾驶
- 虚拟现实：游戏，家居装修，旅游，教育，网购，医疗.....
- 增强现实：教育，家居装修
- 人机交互：体育训练，康复训练，装配仿真.....
- 生活服务：商场购物指南，机场导航，旅游景点自助游.....
- 文化遗产保护
- 数字化军事作战
- 智慧城市，智慧校园、数字地球，国家安全
-

机器人导航，无人驾驶

- 工业4.0和中国制造2025，新一代智能机器人的重要技术

无人驾驶技术是一门涉及传感器、计算机、信息通讯、自动控制、导航定位、机器视觉、人工智能等诸多前沿科学的综合技术。

关键技术	介绍
环境感知	通过多种传感器对车辆周围的环境信息进行感知。环境信息车辆自身状态和周围状态信息。
定位导航	包括定位和导航，是无人驾驶技术能否实现的重要依据。
路径规划	在于为智能车提供最优的行驶路径，包括行驶路线确定、障碍物躲避、转口转向等。
决策控制	相当于智能车的大脑，综合分析环境感知系统提供的信息，对当前的车辆行为产生决策。



Honda(本田)Asimo机器人



Google



百度

虚拟现实

谷歌Cardboard；Oculus Rift；HTC Vive；三星Gear VR；索尼PS VR。

“在场”（Presence）是一个行业术语，用来描述一种VR体验，即让大脑认为自己正处于所见到或正在互动的环境或场景中。



Oculus VR 眼镜



淘宝推出全新购物方式Buy+。
Buy+（败家）使用VirtualReality技术，利用计算机图形系统和辅助传感器，生成可交互的三维购物环境。Buy+将突破时间和空间的限制，真正实现各地商场随便逛，各类商品随便试。

增强现实

- A technology that enhances and extends human vision by fusing images and virtual objects



- 联想和Google合作的全球第一款使用Project Tango技术的AR手机正式发布，产品名为Phab2 Pro

举起手机就能让恐龙站在你面前，而且画面可以根据你的位置来变化，所以你可以走近它甚至绕过另一面来观看。

-



生活服务

精细导航到某个楼层





虚拟广告



自助旅游，哪里不懂照那里

中科院自动化所、模式识别国家重点实验室、机器人视觉组：
<http://vision.ia.ac.cn/>

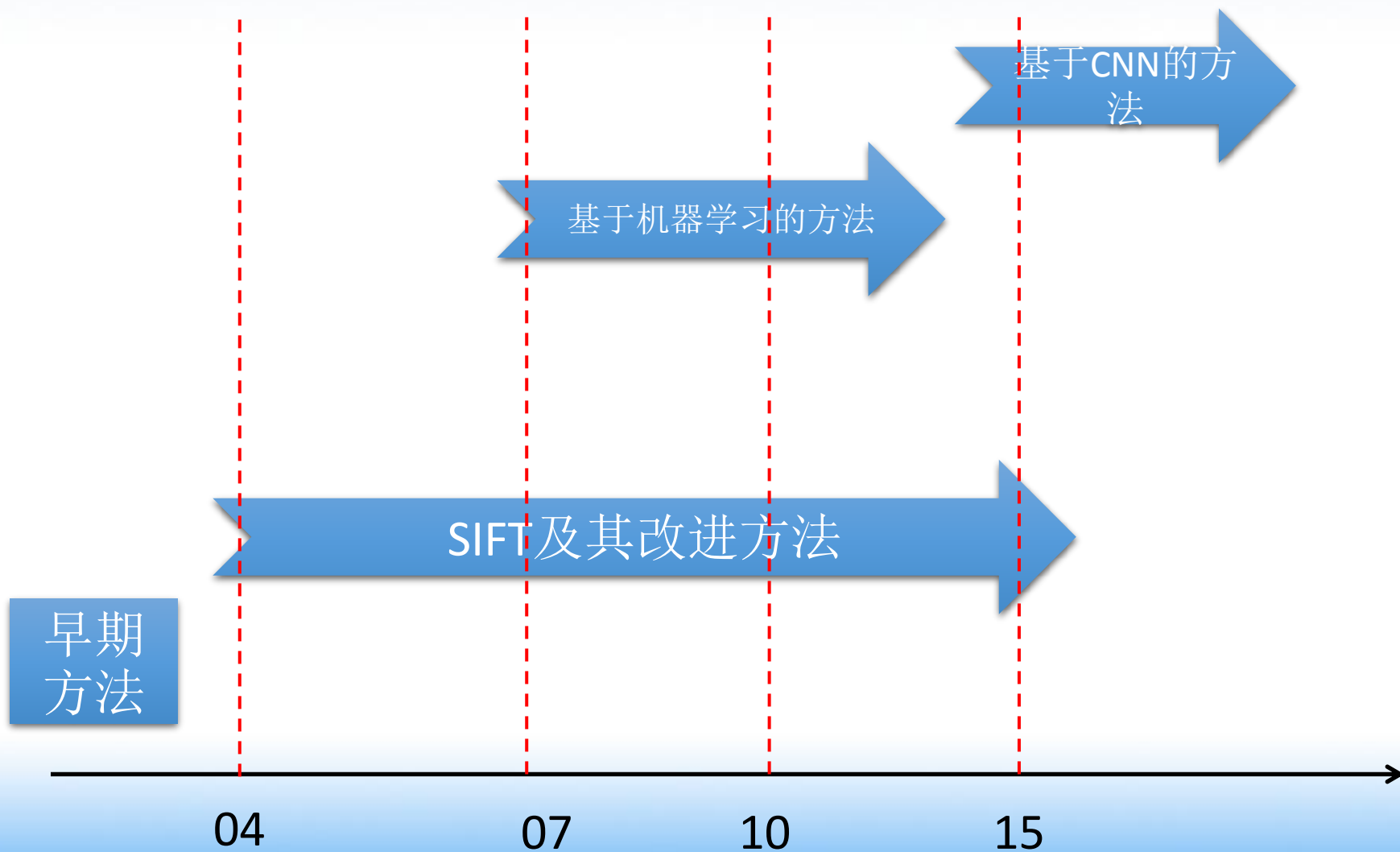
专注三维计算机视觉研究20余载：

多视几何理论、三维视觉认知、几何学习、
图像拟合、图像匹配、
相机标定、相机定位、三维重建

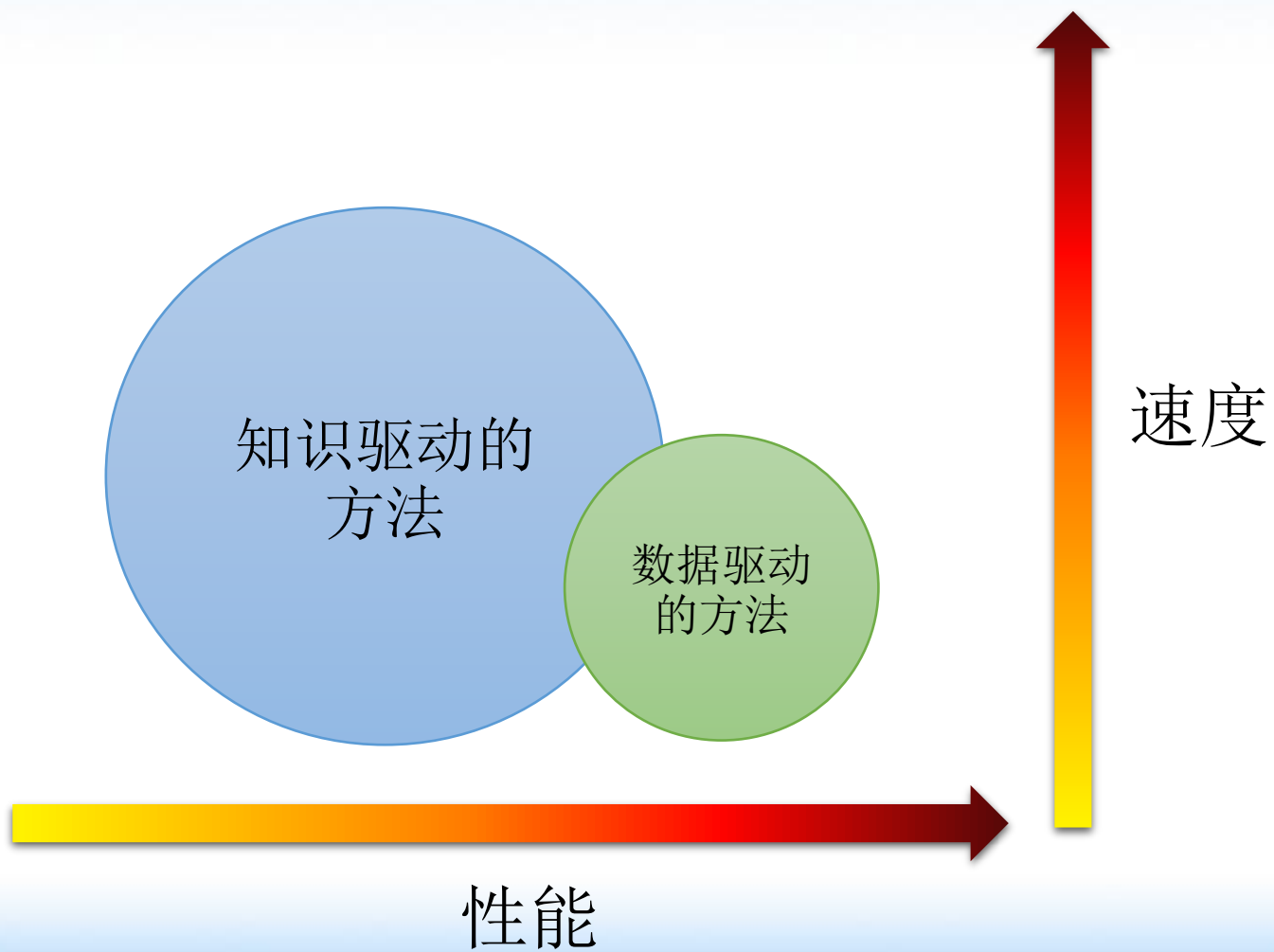
应用：SLAM（ipad、手机端、PC端）、桌面环境快速重建、
大规模城市场景重建、军事地形重建、古建筑重建、考古
现场重建（手持设备、无人机、街景车）.....

图像匹配

图像描述子发展趋势



性能概述



[1.知识驱动的方法]特征描述的灰度序理论

特征汇聚的灰度序方法

底层特征计算的灰度序方法

**LIOP, MROGH, MRRID,
OIOP, MIOP**

ICCV2011, CVPR2011,
PAMI2012, ECCV2014,
PAMI2016

[2.数据驱动的方法]深度神经网络特征学习

面向欧式空间
匹配特性的特征学习

引入中间卷积层的监督信息
提高泛化能力

L2 Net

ECCV2016 竞赛
CVPR2017

在Krystian Mikolajczyk最新的性能评测(文章发表在ICPR' 12)中, 提出的灰度序特征描述子 (**MROGH**、**MRRID**、**LIOP**), 表现出**最好的**匹配性能。

Evaluation of Local Detectors and Descriptors for Fast Feature Matching

Ondrej Miksik

CMP, Czech Republic

ondra.miksik@gmail.com

Krystian Mikolajczyk

CVSSP, UK

k.mikolajczyk@surrey.ac.uk

4 Conclusion

This paper reports the matching precision and speed of recent feature detectors and descriptors in multiple randomized kd-tree setup and Hamming distance brute-force search. The recently proposed real valued descriptors such as **LIOP**, **MRRID**, **MROGH** outperform state-of-the-art descriptors SURF and SIFT in both, precision and recall, although their efficiency is very low. Similarly, binary descriptors provide a very efficient technique for time-constrained applications with good matching accuracy. Another advantages of binary descriptors are very fast extraction times and very low

Detector	Descriptor	Precision	Recall	MAP
SURF	SURF	0.485	0.513	0.334
SURF	SIFT	0.525	0.533	0.491
SURF	BRIEF	0.517	0.546	0.514
SURF	ORB	0.448	0.470	0.437
SURF	LIOP	0.581	0.597	0.568
SURF	MROGH	0.540	0.567	0.527
SURF	MRRID	0.550	0.569	0.510
SURF	BRISK	0.536	0.553	0.530
BRISK	BRISK	0.504	0.527	0.492
ORB	ORB	0.493	0.495	0.463
FAST	SIFT	0.366	0.376	0.336

注: Krystian Mikolajczyk 的上一个评测工作 (2005年完成) 获得了**CVPR 2014 Longuet-Higgins Prize**。

PAMI最近接收的一篇文章里，评价灰度序方法代表了旋转不变描述子的state of the art，并且在Oxford数据集上取得了最好的效果

📄 Rethinking the sGLOH Descriptor

Fabio Bellavia; Carlo Colombo

IEEE Transactions on Pattern Analysis and Machine Intelligence

Year: 2017, Volume: PP, Issue: 99

Pages: 1 - 1, DOI: 10.1109/TPAMI.2017.2697849

IEEE Early Access Articles

include the well known SIFT, that is considered as reference, LIOP, MIOP and MROGH, that represent the state-of-the-art for rotational invariant descriptors, and

based strategies, using fast matching only slightly lowers the performances achieved with full matching. Fully rotational invariant LIOP and MROGH achieve the best results, followed by sGLOH2 and BisGLOH2 strategies,

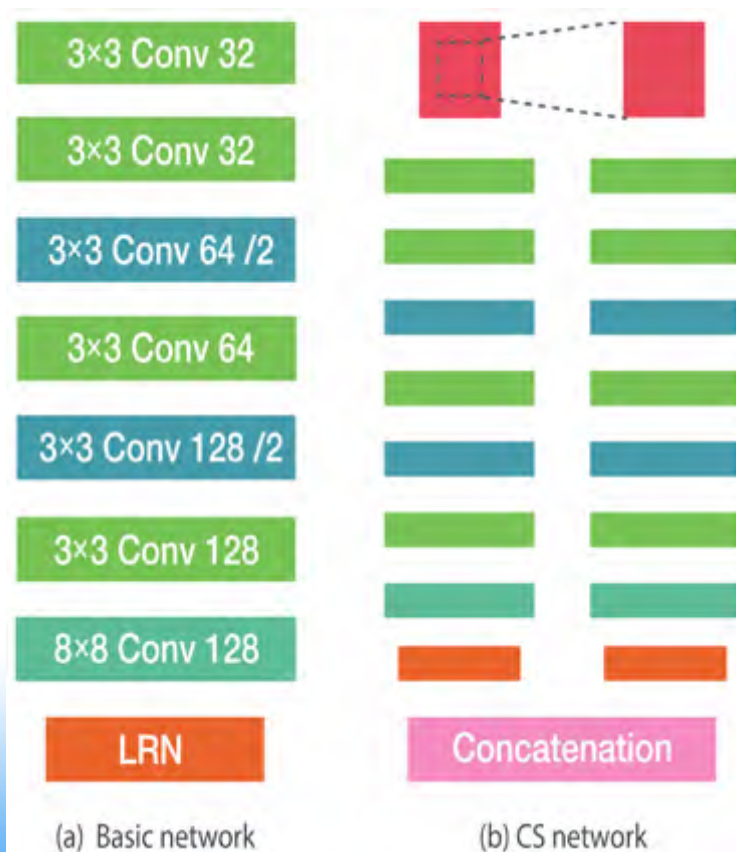
RESULTS ON THE OXFORD DATASET. LOWER VALUES ARE BETTER FOR MSRE AND MSAP, HIGHER FOR MAP. SEE TEXT FOR DETAILS

	msRE (%)				msAP (%)				mAP (%)			
	Histogram		Binary		Histogram		Binary		Histogram		Binary	
	Std	Fast	Std	Fast	Std	Fast	Std	Fast	Std	Fast	Std	Fast
SIFT	6.1	—	—	—	7.4	—	—	—	60.3	—	—	—
VL SIFT	5.6	—	—	—	5.7	—	—	—	58.0	—	—	—
VL SIFT*	5.6	—	—	—	5.5	—	—	—	59.9	—	—	—
LIOP	2.3	—	—	—	2.3	—	—	—	69.5	—	—	—
MIOP	1.5	—	—	—	1.8	—	—	—	72.6	—	—	—
MROGH	0.4	—	—	—	0.5	—	—	—	75.6	—	—	—
DeepDesc	8.8	—	—	—	9.5	—	—	—	48.4	—	—	—
RFD _r	—	—	5.4	—	—	—	5.5	—	—	—	61.6	—
RFD _g	—	—	5.2	—	—	—	5.2	—	—	—	62.5	—
sGLOH	3.4	3.8	4.5	4.7	3.3	3.7	4.2	4.4	62.4	61.1	57.9	57.4
sGO _r	2.5	3.1	3.5	3.9	2.4	3.1	3.4	3.8	65.3	62.9	61.0	59.9
sGLOH2	2.5	2.8	3.6	3.8	2.5	2.7	3.4	3.6	68.2	67.0	64.0	63.6
sGO _r 2 _a	1.6	2.0	2.3	2.7	1.5	2.0	2.2	2.5	71.2	69.9	68.8	67.7
sGO _r 2 _h	1.6	2.0	2.4	2.7	1.5	2.0	2.2	2.6	71.4	69.1	67.9	66.8

深度学习特征描述方法, L2-Net

现有方法的**局限性**:

- 泛化能力**弱**、普适性**不强**
- 与度量学习耦合，匹配过程**复杂**、速度**慢**



- Input: 32*32 patche
- Output: 128位浮点型描述子

田雨润 樊斌 吴福朝
ECCV 2016 描述子竞赛三项指标第一
CVPR2017

网络结构:

6个 3×3 卷积层;

1个 8×8 卷积层;

CIC (Convolution in Convolution) 只作用于训练阶段的误差层, 限制匹配的patch有相似的特征, 不匹配的patch有不同的特征

LRN (Local Response Normalization) 将网络输出归一化

训练

多误差约束:

相似性误差: 相对距离

紧致性误差: 降低**128**位描述子位与位之间相关性, 防止过拟合

泛化能力误差: 为增强泛化能力, 误差加在中间层, 无需更多参数

2至4小时, GPU

描述子提取速度:

21.3 k patch/s

The proposed network is named as L2-Net since the output descriptor can be matched in Euclidean space by L2 distance

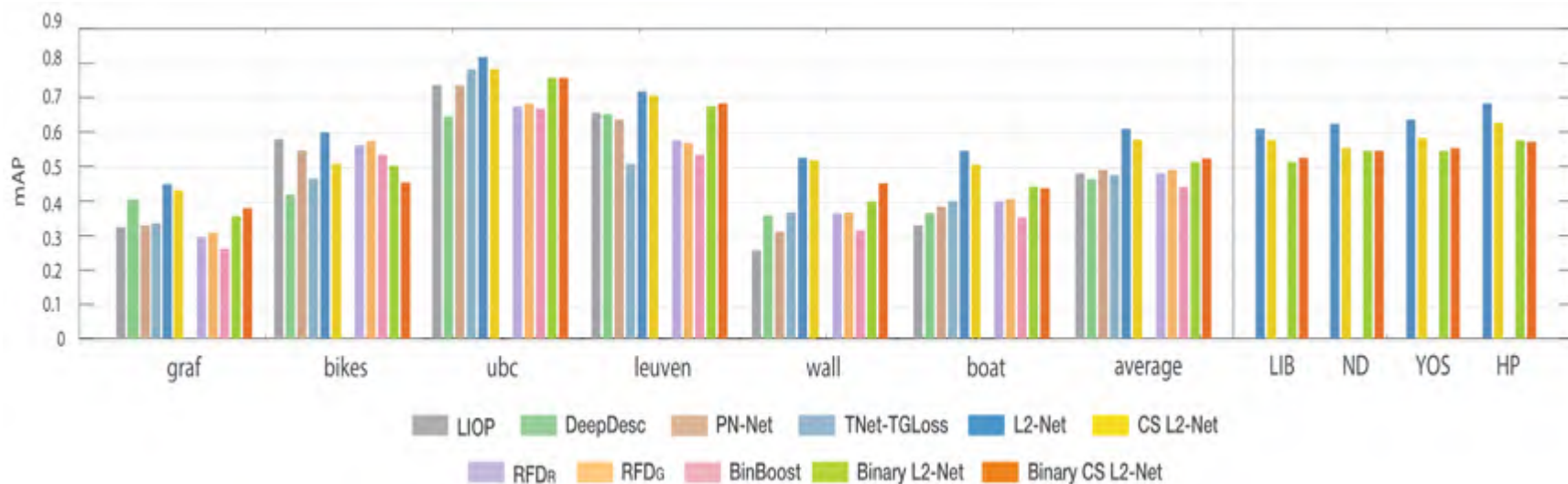
Performance on Brown dataset

Training Test	Feature Dim	Notredame		Yosemite		Liberty		Mean
		Liberty	Notredame	Yosemite	Liberty	Notredame		
Metric Learning								
SIFT [15]	128	29.84		22.53		27.29		26.55
MatchNet [10]	4096	6.9	10.77	3.87	5.67	10.88	8.39	7.74
DeepCompare 2ch-2stream [25] +	256	4.85	7.20	1.90	2.11	5.00	4.10	4.19
DeepCompare 2ch-deep [25] +	256	4.55	7.40	2.01	2.52	4.75	4.38	4.26
SNet-GLoss [9] +	256	6.39	8.43	1.84	2.83	6.61	5.57	5.27
CS SNet-GLoss [9] +	384	3.69	4.91	0.77	1.14	3.09	2.67	2.71
Float Descriptors								
TNet-TGLoss [9] +	256	9.91	13.45	3.91	5.43	10.65	9.47	8.8
TNet-TLoss [9] +	256	10.77	13.90	4.47	5.58	11.82	10.96	9.58
PN-Net [2]	256	8.13	9.65	3.71	4.23	8.99	7.21	6.98
DeepDesc [5]	128	10.9		4.40		5.69		6.99
L2-Net	128	3.64	5.29	1.15	1.62	4.43	3.30	3.23
L2-Net +	128	2.36	4.7	0.72	1.29	2.57	1.71	2.22
CS L2-Net	256	2.55	4.24	0.87	1.39	3.81	2.84	2.61
CS L2-Net +	256	1.71	3.87	0.56	1.09	2.07	1.3	1.76
Binary Descriptors								
RFD _R [7]	293-598	19.35	19.40	13.23	11.68	16.99	14.50	15.85
RFD _G [7]	406-563	17.77	19.03	12.49	11.37	17.62	14.14	15.4
BinBoost [21]	64	20.49	21.67	16.90	14.54	22.88	18.97	19.24
RMGD [26]	1376-1600	15.09	17.42	10.15	10.86	14.46	13.82	13.63
Boixet al [29]	1360	15.6	15.52	-	8.52	-	8.87	12.12
Binary L2-Net	128	10.3	11.71	6.37	6.76	13.5	11.57	10.03
Binary L2-Net +	128	7.44	10.29	3.81	4.31	8.81	7.45	7.01
Binary CS L2-Net	256	5.25	7.83	3.07	3.52	8.49	6.92	5.84
Binary CS L2-Net +	256	4.01	6.65	1.9	2.51	5.61	4.04	4.12

- ◆ Accompany with the float L2-Net descriptor, we further obtain a binary descriptor by simply taking the sign of the float descriptor (± 1). The resulting binary descriptors are denoted as Binary L2-Net and Binary CS L2-Net.

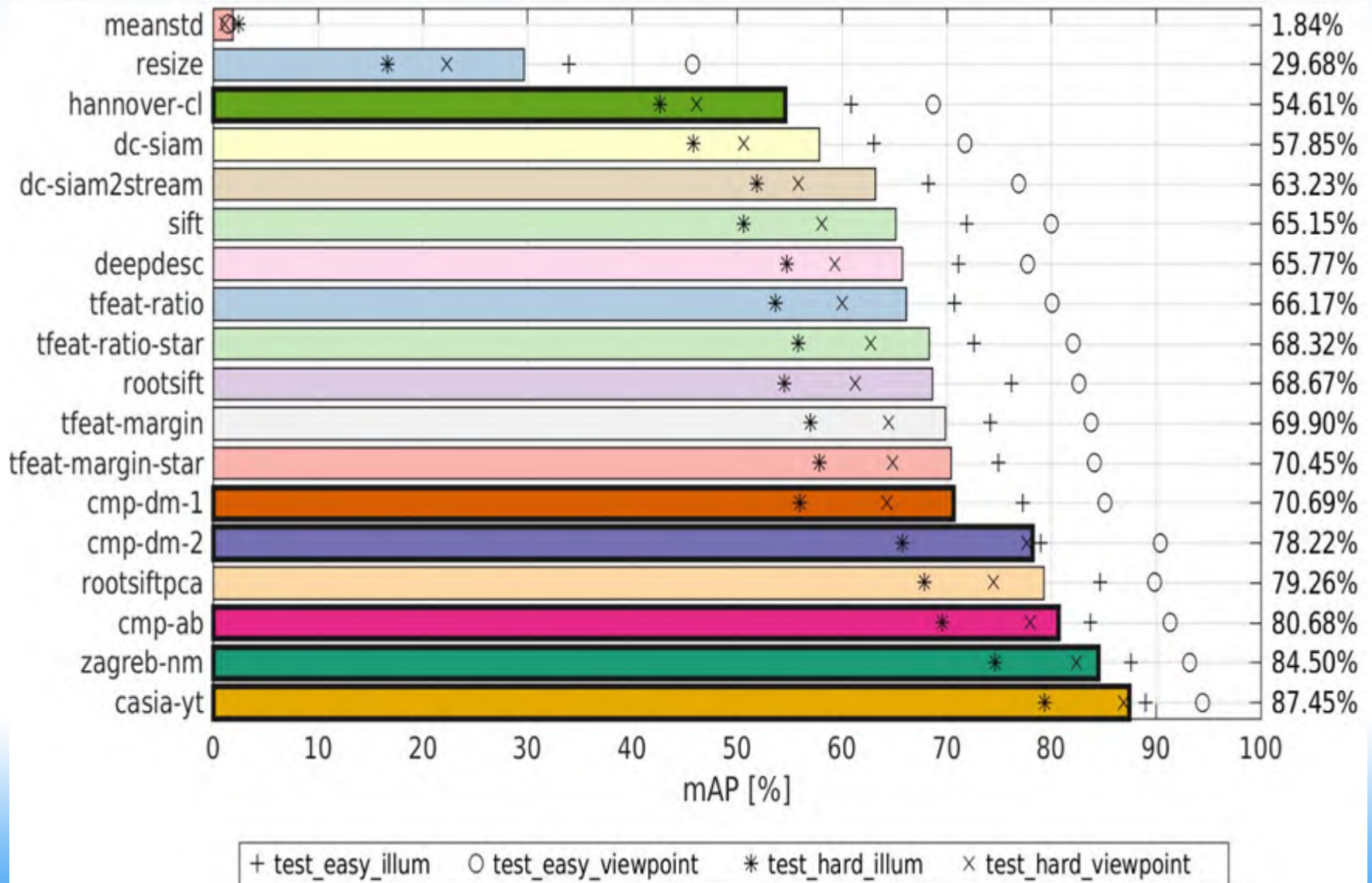
Table 1. Performance on the Brown dataset. The numbers are false positive rate at 95% recall. + indicates data augmentation.

Performance on Brown dataset

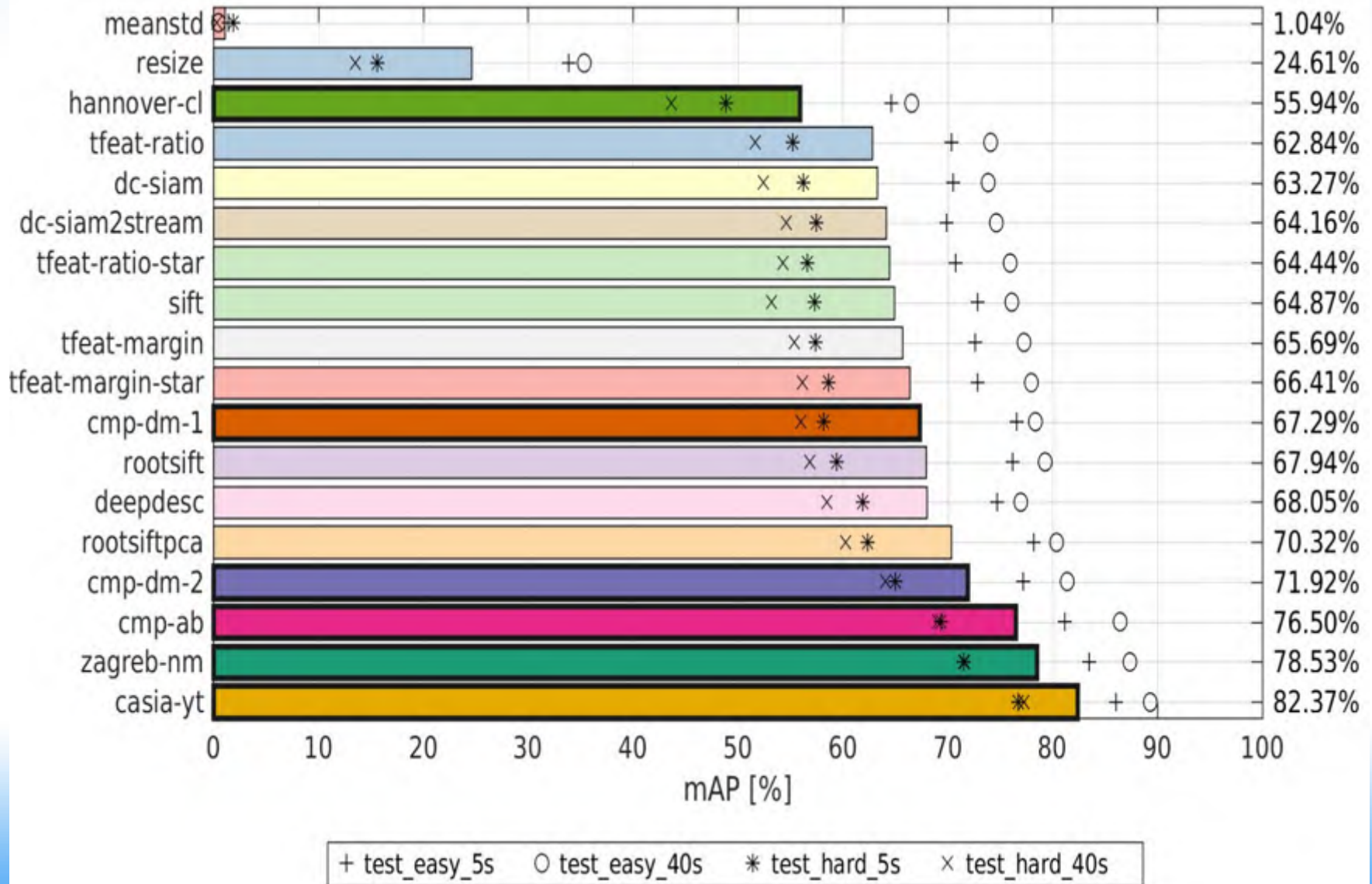


Performance on Hpatches dataset

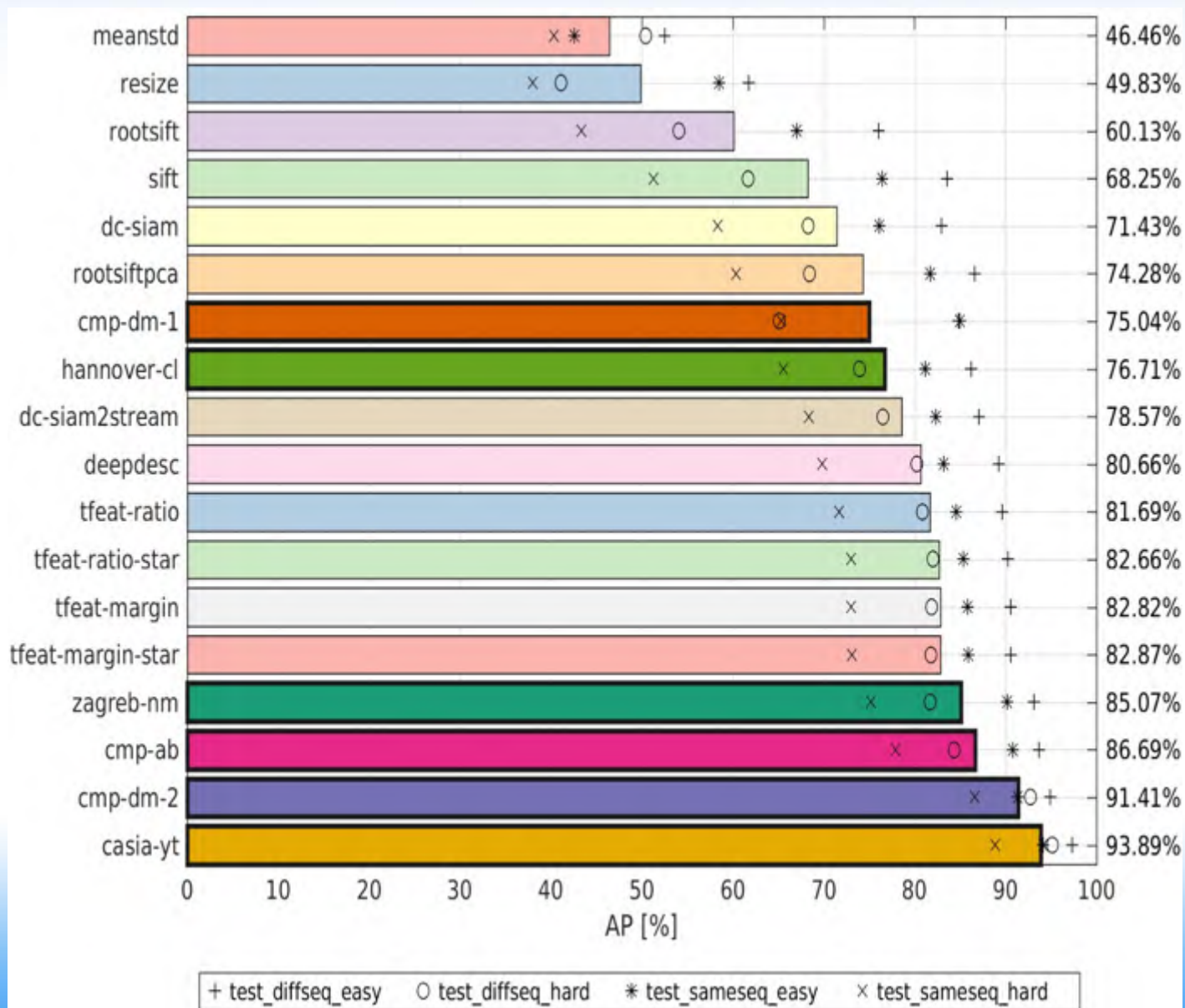
Image matching



Patch retrieval



Patch classification



视觉定位

SLAM(Simultaneous Localization and Mapping)

主要解决的是相机在空间中的定位、以及创建环境的地图

起源1986年，于1996 International Symposium on Robotics Research中正式命名。

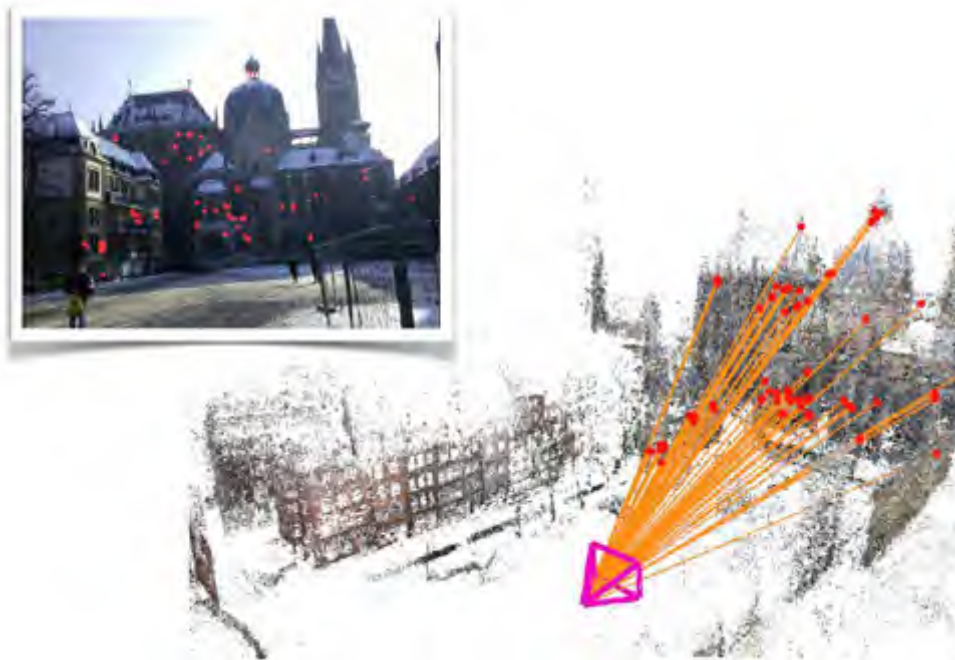
若是手机离开了WIFI和数据网络，就像无人车和机器人，离开了SLAM一样。

SLAM 重定位

- 重定位的必要性
 - sudden illumination change
 - full occlusion
 - extreme motion blur
- 任务
 - re-estimating the camera pose after tracking failure

SLAM 重定位

- 基于随机森林的城市大数据（千万级）的毫秒级定位



- 基于Local Sensitive Hashing的大数据实时定位

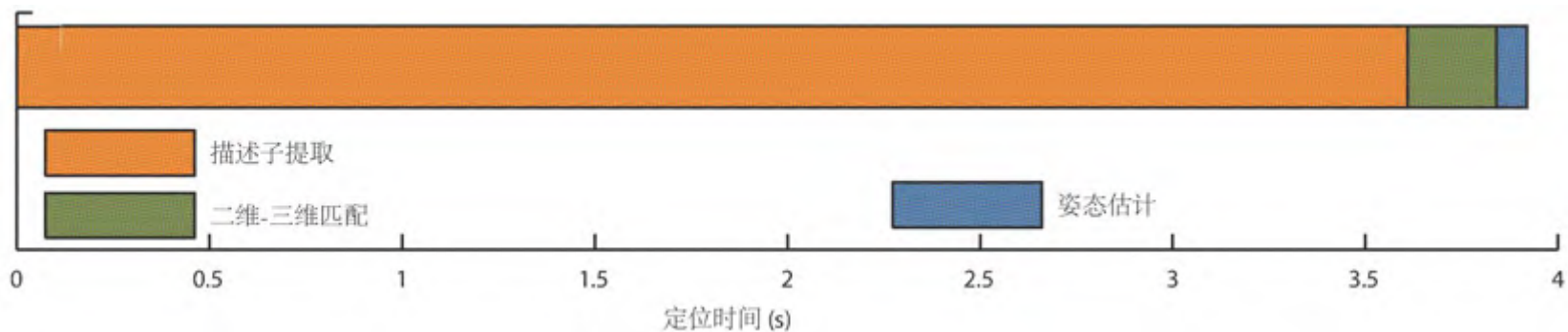
冯友计 吴毅红 范力欣

IEEE Trans. On Image Processing 2016;

Machine Vision Applications 2017

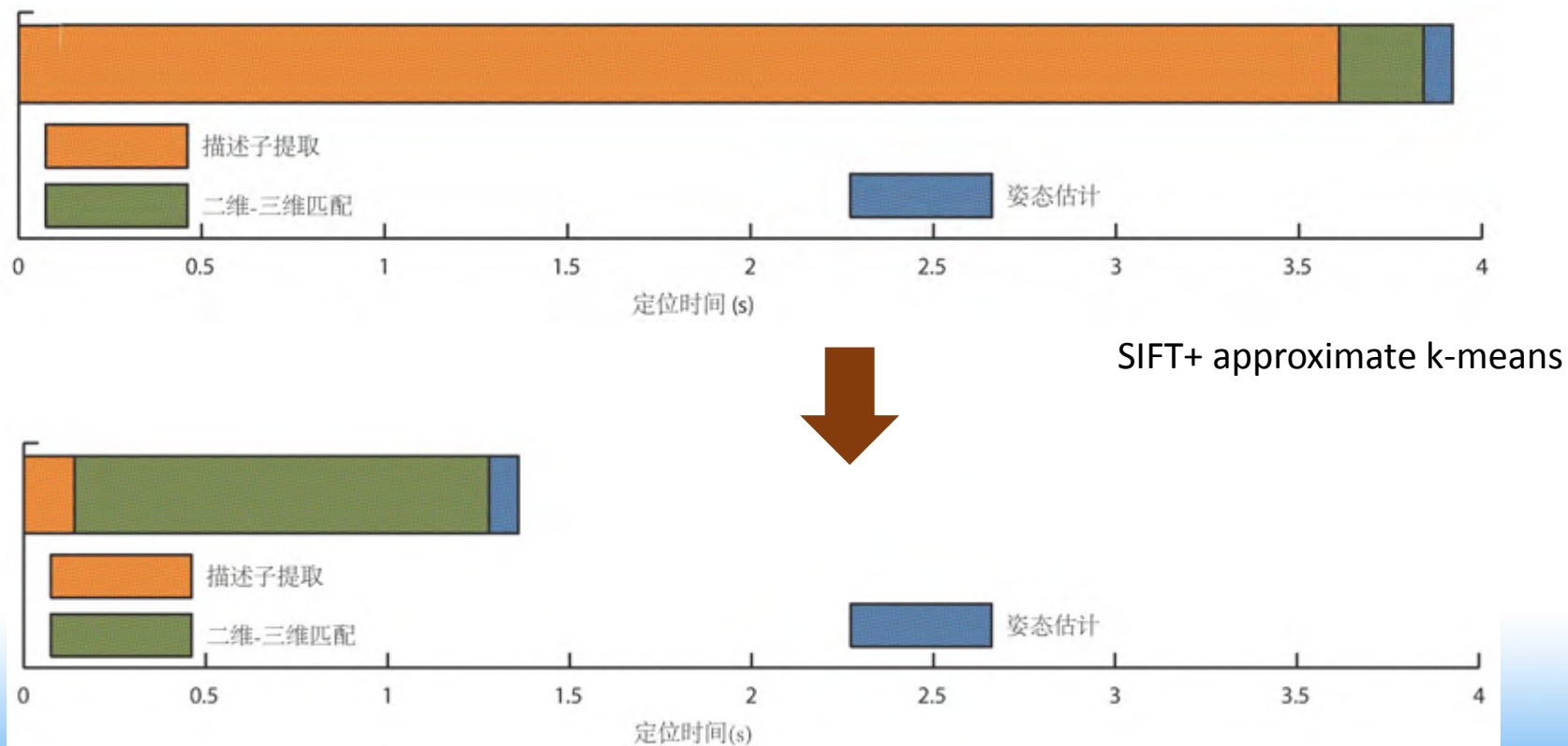
- **The problem**

As can be seen, the descriptor extraction, which takes 3.61s, almost accounts for half of the localization time



• One direct way

- Binary descriptors such as BRISK, ORB, FRIF to substitute SIFT.



基于随机森林的城市大数据毫秒级定位

- We construct randomized trees in a supervised training process by exploiting label information:
 - The existing indexing methods: unsupervised ignoring the label information.
 - The cost function:

$$C = \lambda(1 - r_c) + \frac{1}{0.5 - |0.5 - \mu|}$$

error rates uniform leaf sizes

where $\mu = \frac{|S_l|}{(|S_l| + |S_r|)}$ and $|\cdot|$ is the size of the set. λ is used to place a trade off between the accuracy and the uniform.

- We propose an indexing approach for approximate nearest neighbor search of binary features: to priorly search the leaf node with the largest success probability.

Experiments

- On retrieval
- On localization

Setup: a laptop with 16GB RAM and an Intel Core i3-2310
2.1GHz CPU

On retrieval

Dataset: Dubrovnik, Rome, and Aachen

- 1,500,000 3D points
- 5,270,000 database features
- 1,800,000 query features

Compared approaches

- RT_AP: the randomized trees with supervised training and with priority search
- RT: the randomized trees with supervised training and without priority search
- RT_b: the randomized trees without supervised training and without priority search
- LSH: locality sensitive hashing
- MLSH: multi-probe locality sensitive hashing
- HCT: hierarchical clustering trees
- EWH: error weighted hashing
- MIH: multi-indexing hashing

On retrieval

RT_AP	(0.45, 0.02)	(0.58, 0.07)	(0.67, 0.23)
RT	(0.46, 0.03)	(0.56, 0.18)	-
RT_b	(0.44, 0.08)	-	-
HCT_16	(0.45, 0.10)	(0.55, 0.22)	-
HCT_32	(0.43, 0.11)	(0.58, 0.28)	(0.66, 0.78)
LSH	(0.46, 0.18)	(0.56, 1.06)	-
MLSH_1	-	(0.57, 0.48)	-
MLSH_2	-	(0.57, 0.62)	(0.67, 2.16)

The efficiency of the proposed RT_AP is remarkably higher than LSH, MLSH and HCT. Under similar retrieval accuracies,

- 3 to 4 times faster than HCT 32 and HCT 16,
- 9 to 14 times faster than LSH,
- 6 times faster than MLSH 1
- 8 times faster than MLSH 2.

On localization

- Dataset

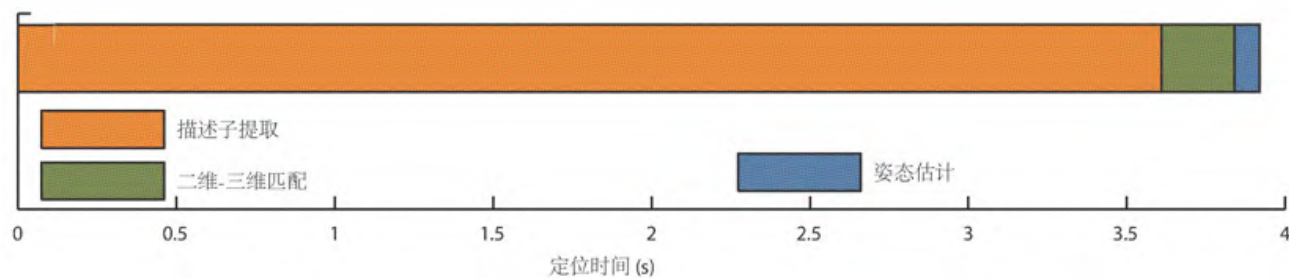


数据集	数据库中 图像数目	数据库中 三维点数目	数据库中 特征数目	查询 图像数目	单幅查询图像 平均特征数目
Dubrovnik	6044	1,886,884	9,606,317	800	9768
Rome	16,179	4,312,820	21,515,110	1000	7279
Aachen	3047	1,540,786	7,281,501	369	8648

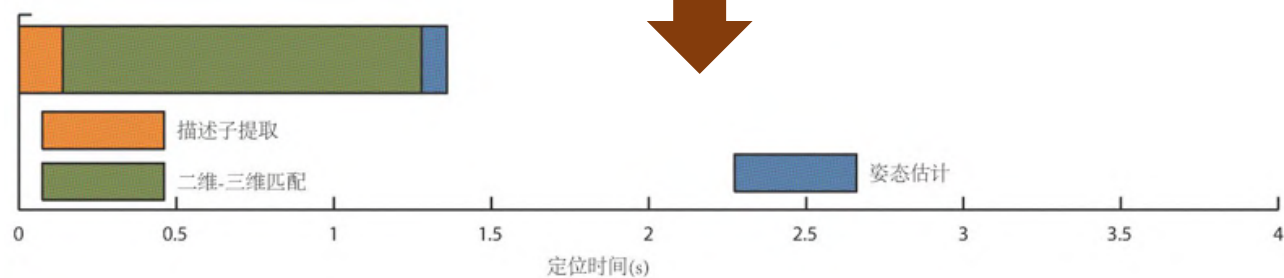
On localization

Dataset	Method	# registered images	descriptor extraction [s]	2D-3D correspondence search [s]	RANSAC [s]	total [s]
Dubrovnik	DM [7]	780.4	3.61	3.97+0.43	0.05	8.06
	P2F [6]	753	3.61	-	-	4.34
	RT_AP	784.9	0.13	0.18	0.10	0.42
	RT_AP(G)	781.6	0.13	0.18	0.11	0.43
Rome	DM [7]	972.7	2.69	2.98+0.61	0.02	6.30
	P2F [6]	921	2.69	-	-	3.60
	RT_AP	978.8	0.11	0.14	0.04	0.29
	RT_AP(G)	979.7	0.11	0.21	0.04	0.36
Aachen	DM [7]	295.4	3.20	3.54+0.47	0.06	7.27
	RT_AP	301.2	0.12	0.18	0.08	0.37
	RT_AP(G)	299.9	0.12	0.20	0.10	0.42

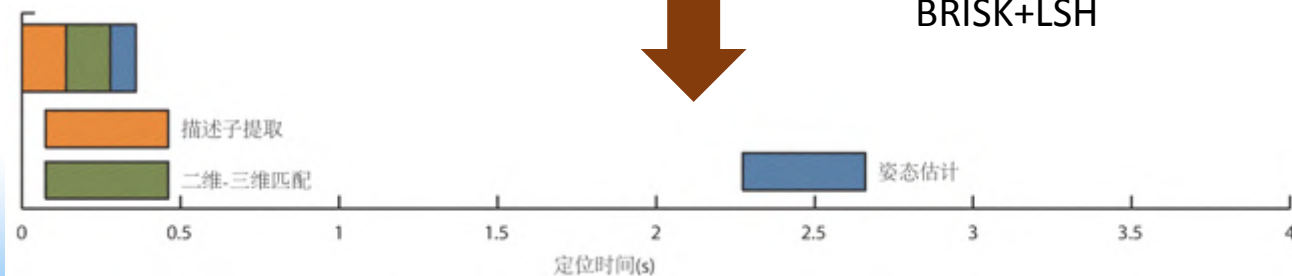
On localization



SIFT+ approximate k-means



BRISK+LSH



BRISK+RT_AP

On localization

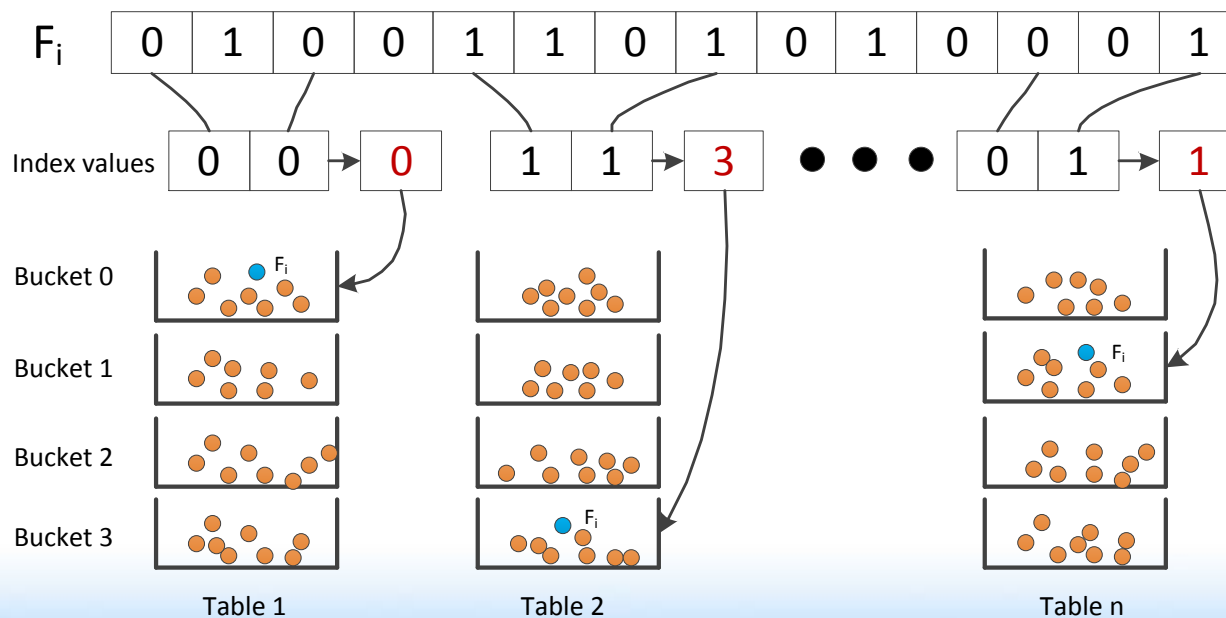
- Fast Camera Relocalization in SLAM



基于Local Sensitive Hashing的实时重定位

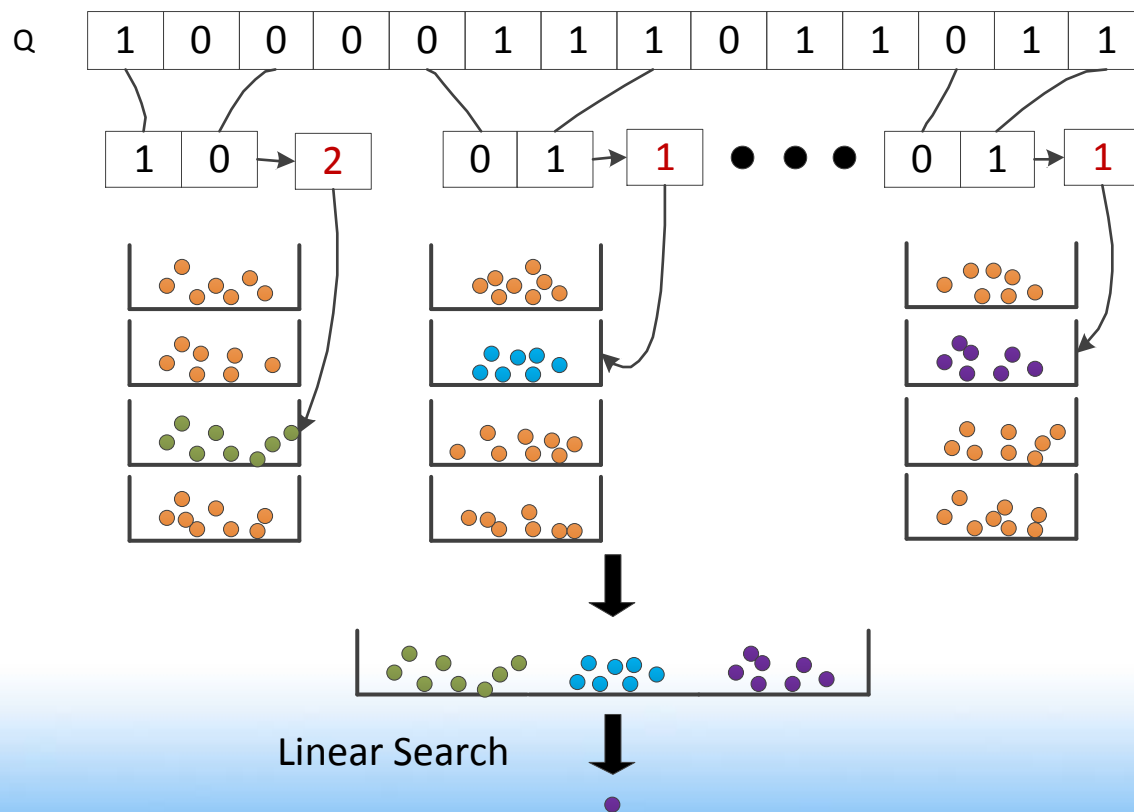
- 采用多个哈希键索引特征
- 随机选取特征的某些比特位构成哈希键

$$\mathbb{H}^1 = \{b_1, b_3\}; \mathbb{H}^2 = \{b_5, b_8\}; \mathbb{H}^n = \{b_{12}, b_{14}\}$$



数据库特征 F_i 的索引和存储过程

- 获取查询特征对应的哈希桶
- 在哈希桶中线性搜索



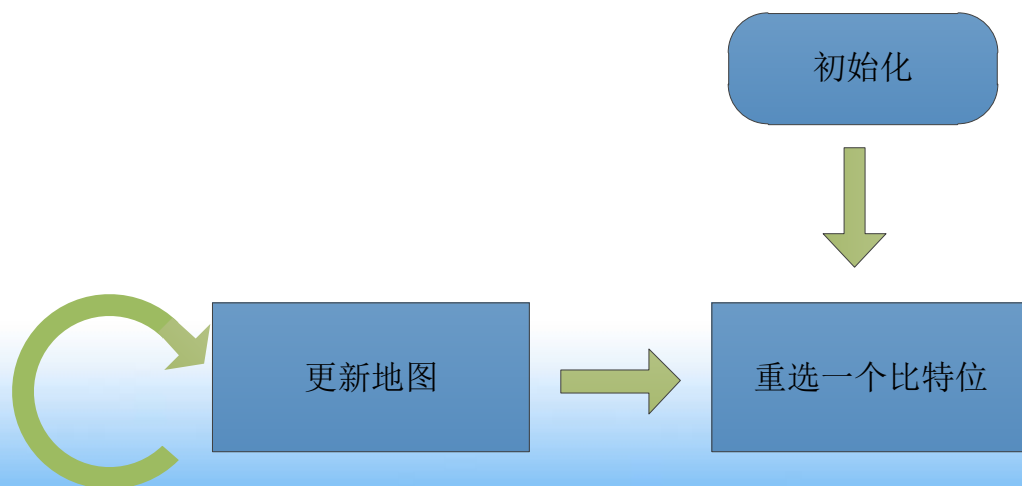
查询特征 Q 的近似最近邻搜索

搜索效率的提升

- 影响搜索效率的因素
 - 哈希桶大小（存储的数据库特征的数量）的均匀性
 - 相互匹配的特征的哈希碰撞率，即相互匹配的特征出现在同一个哈希桶中的概率
- 提出的方法
 - 通过学习构造哈希键来获得
 - 大小尽量均匀的哈希桶
 - 尽量高的哈希碰撞率

哈希键的构造

- 初始化
 - 随机选取特征的某些比特，与原始的LSH相同
- 在线学习
 - 重新选择构成哈希键的比特位来最小化关于均匀性和碰撞率的损失函数
 - 只在添加关键帧，并且更新地图时进行，每次重选哈希键的一个比特



损失函数

$$C = \lambda(1 - p_c) + \frac{1}{1 - u'}$$

碰撞率

$$u' = \frac{\sum_{n=1}^N s_n^2}{\sum_{m=1}^{N/2} \tilde{s}_m^2}$$

其中N是哈希桶的数量，

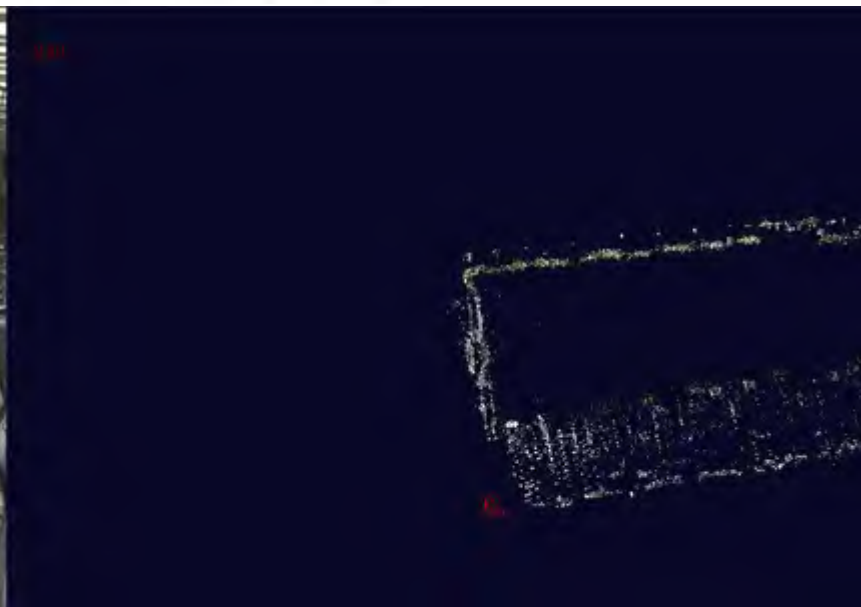
s_n 是第n个归一化后哈希桶的大小

$\{\tilde{s}_m, m = 1, 2, \dots, \frac{N}{2}\}$ 是用未改变的哈希键索引数据库特征后各哈希桶的大小。

实时性:

视频段	V1		V2	
查询帧数	8000		1063	
三维点数	37,641		4173	
数据库特征数	175,207		20,846	
索引方法	Original LSH	Learned LSH	Original LSH	Learned LSH
成功重定位帧数	7832.0	7859.0	927.8	953.2
时间(ms)				
角点提取	8.3	8.5	7.8	7.8
描述子提取	8.3	8.3	10.0	10.0
近似最近邻搜索	30.4	17.5	3.7	3.5
RANSAC	0.5	0.5	1.2	1.2
总时间	48.3	35.9	24.8	24.6

- 通过在线学习构造哈希键，提高近似最近邻搜索的效率
- 学习被分散在SLAM的过程中，不会给SLAM系统带来延迟

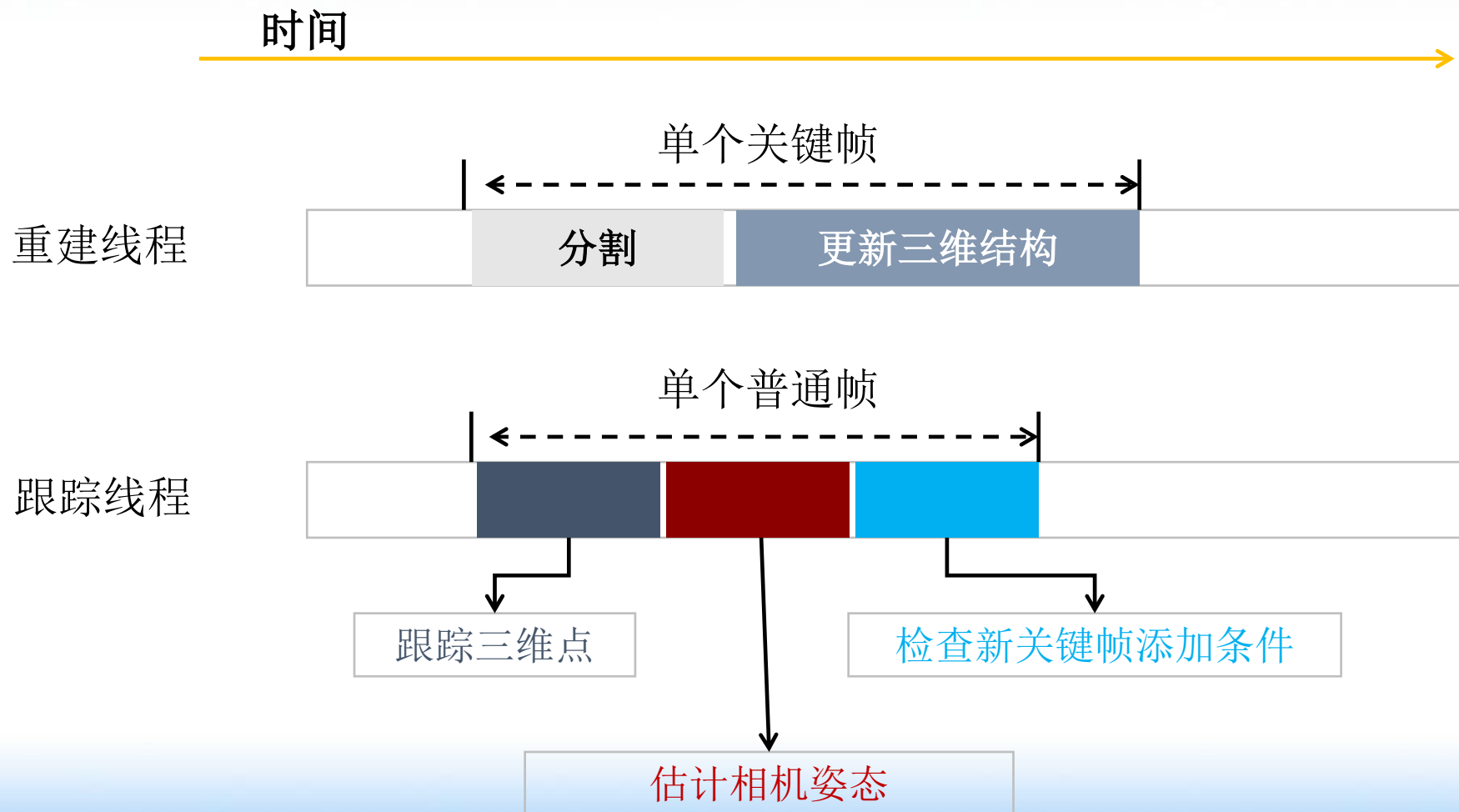


Optical Flow SLAM

- 动态目标在线实时SLAM
- 场景SLAM

优势:

- 快速
- 少纹理也可使用
- 快速运动 鲁棒



- 图像提取：改进Lucas-Kanade Tracker

从初始位置出发，迭代搜寻源图像块（最近的关键帧中三维点的观测周围的图像块）在当前帧中的对应即目标图像块

特点

- 在物体纹理较稀少或者有运动模糊时比先提特征点再匹配的方法更稳定
- 用预估的姿态对源图像块进行仿射变形，可处理尺度、旋转和视角变化
- 对图像块归一化，可处理线性光照变化
- 亚像素级的特征提取，更密集地特征提取

- 重建质量有效控制:

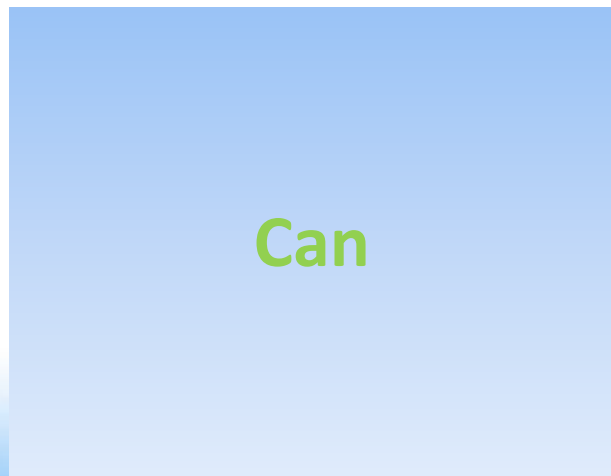
$$Q(\mathbf{X}_i) = \frac{\sqrt{\text{tr}(\Sigma_{\mathbf{X}_i})}}{d_i} \quad \Sigma_{\mathbf{X}_i} = \left(\left(\frac{\partial \mathbf{f}}{\partial \mathbf{X}} \Big|_{\mathbf{X}_i} \right)^T \Sigma_{\mathbf{M}_i}^{-1} \frac{\partial \mathbf{f}}{\partial \mathbf{X}} \Big|_{\mathbf{X}_i} \right)^{-1}$$

- 跟踪质量评价，严格剔除外点条件：
 - 在至少三个关键帧中被观测到（新建的三维点除外）
 - 外点观测的比例少于三分之一。观测的重投影误差大于1.5个像素则该观测为外点观测
 - 在至少90%的关键帧中的投影都位于前景区域
- 鲁棒重定位方法：对快速运动造成的模糊、遮挡、光线变化等

动态目标Optical-Flow-SLAM与PTAM相比:

	多纹理			少纹理		
	重建误差	定位位置误差	定位角度误差	重建误差	定位位置误差	定位角度误差
PTAM	0.0054,2.1%	0.015,1.5%	1.07°	0.0071,2.8%	0.023	1.6°
Dynamic Rigid SLAM	0.0020,0.8%	0.007,0.7%	0.41°	0.0022,0.9%	0.013	0.9°

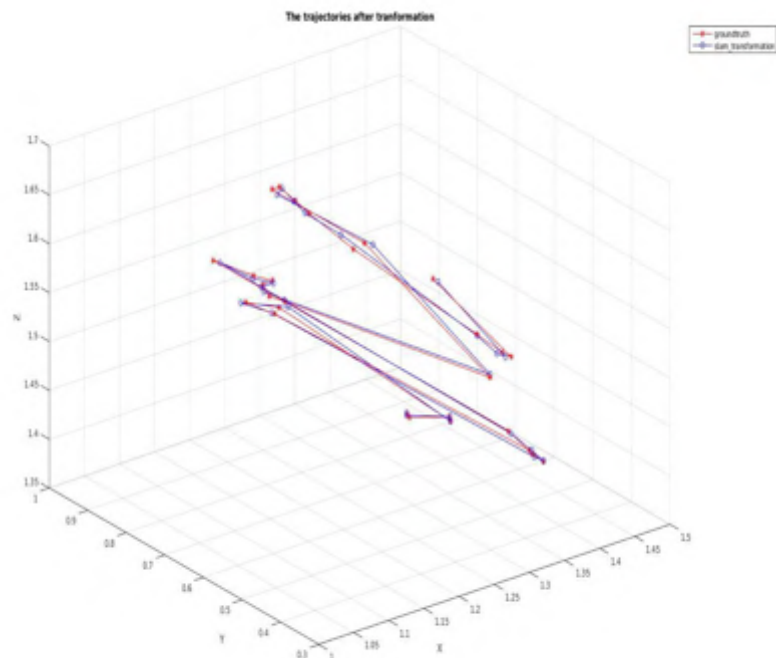
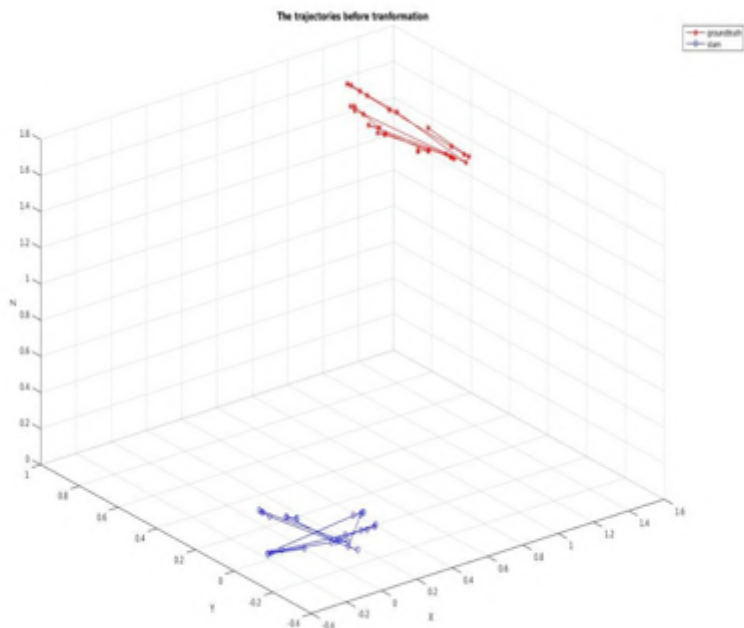
为了适合于PTAM，在视频数据中，只有目标物体，没有任何背景。
PTAM在少纹理的物体上有失败，DR SLAM 都可成功。
PTAM定位有抖动；DR SLAM 更稳定。



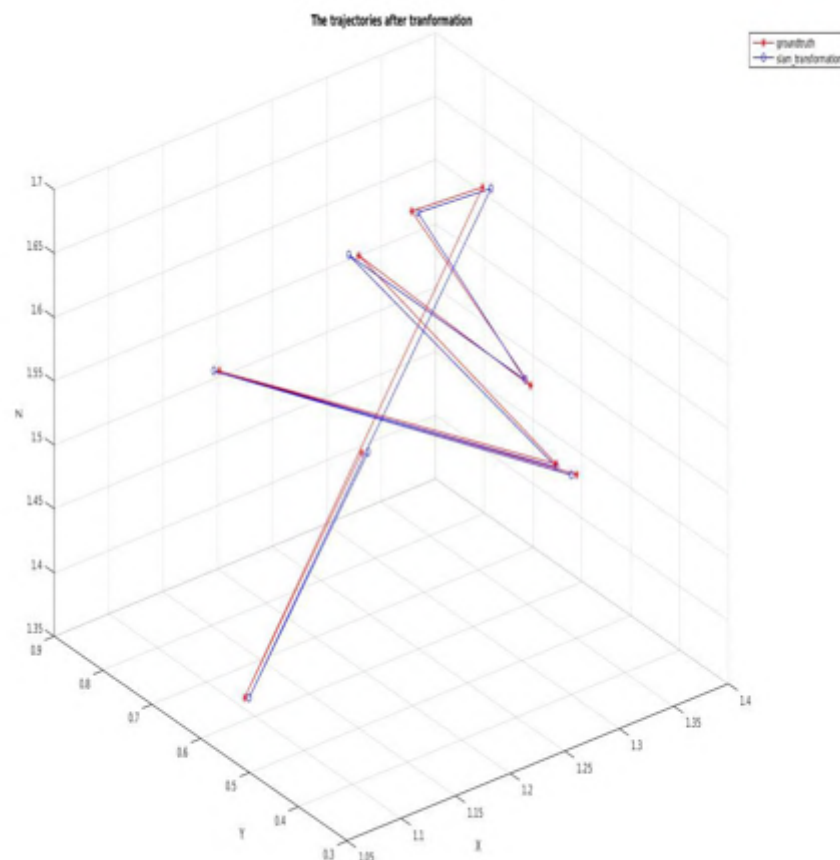
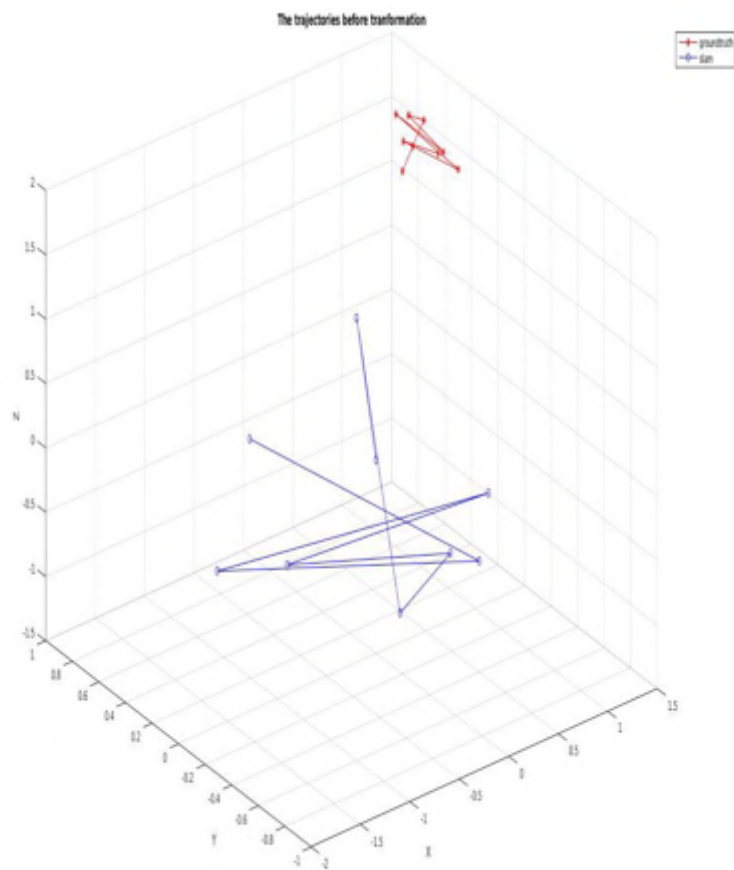
场景Optical-Flow-SLAM与ORB-SLAM 2相比:

fr1_xyz : ORBSLAM(9.5mm)

TUM RGB-D Benchmark



fr1_xyz : OpticalFlowSLAM(7.5mm)

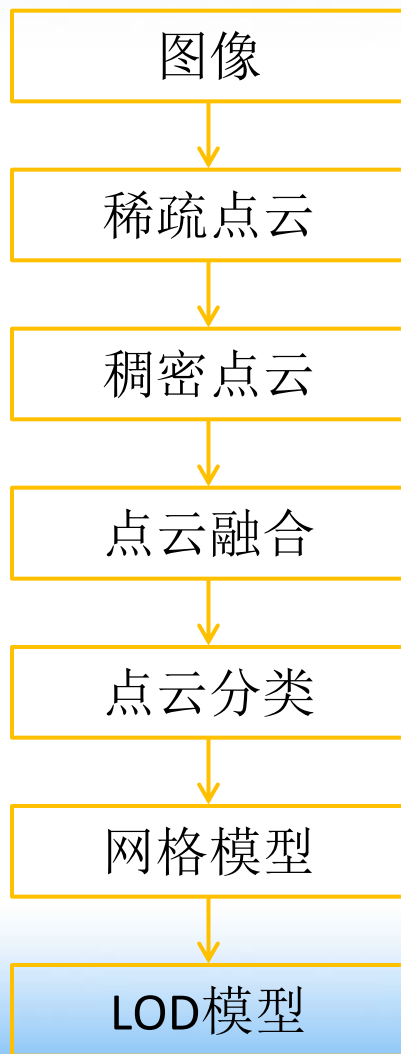


fr1_floor
fr3_nstr_tex_far
fr3_walk_xyz

Optical Flow SLAM

三维重建

图像三维建模系统框架



核心算法:

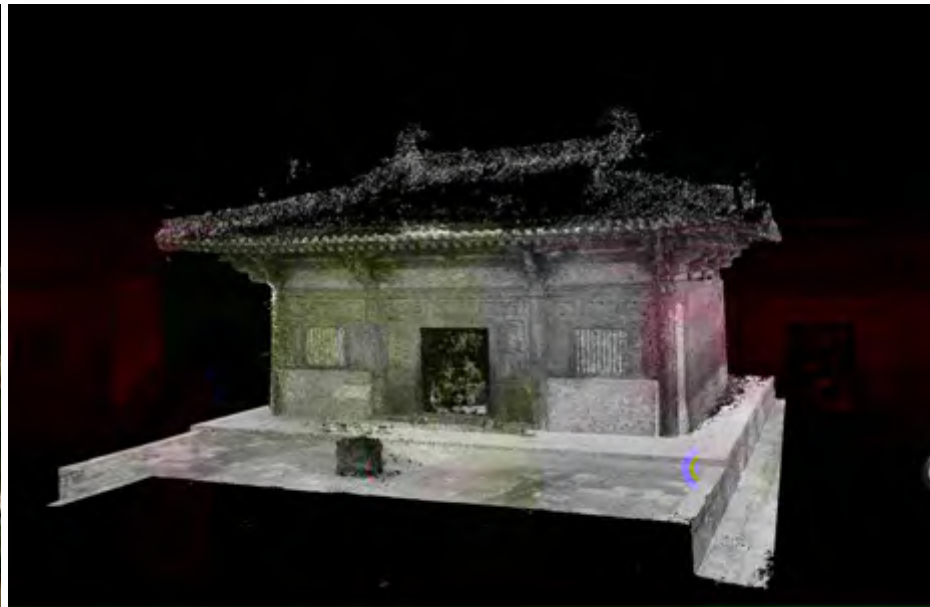
- 1) 融合辅助传感器的稀疏重建
- 2) 全局与增量混合式稀疏重建
- 3) 多相机系统稀疏重建
- 4) 天地图像融合
- 5) 邻域图像组最优选择
- 6) 基于深度图融合的稠密重建
- 7) 点云语义分类
- 8) 点云网格建模与精细化
- 9) 建筑物LOD建模

混合式稀疏重建（无人机图像）

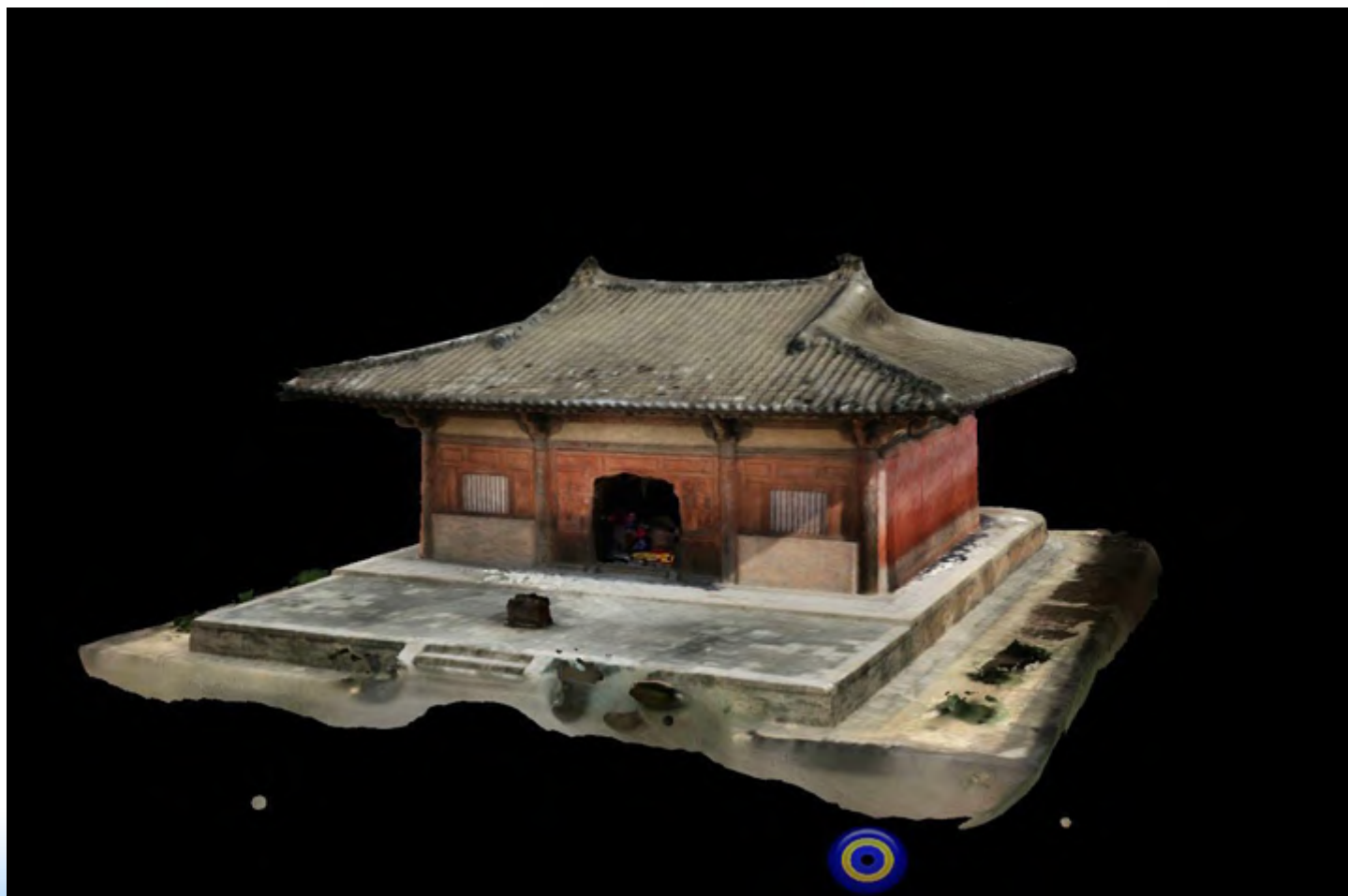


崔海楠，高翔，申抒含，胡占义，CVPR2017，spotlight

无人机图像和地面图像生成的点云融合



天地点云融合及纹理映射后的结果



天地图像融合重建

中国古代建筑三维数字化保护



大同云冈石窟



大同九龙壁



恒山悬空寺



应县木塔



九华旃檀禅林



普陀法雨寺



峨眉金顶



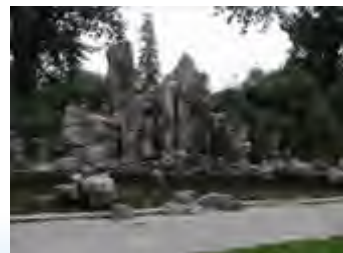
龙泉寺



颐和园石舫



颐和园狮子



清华假山



清华生科楼



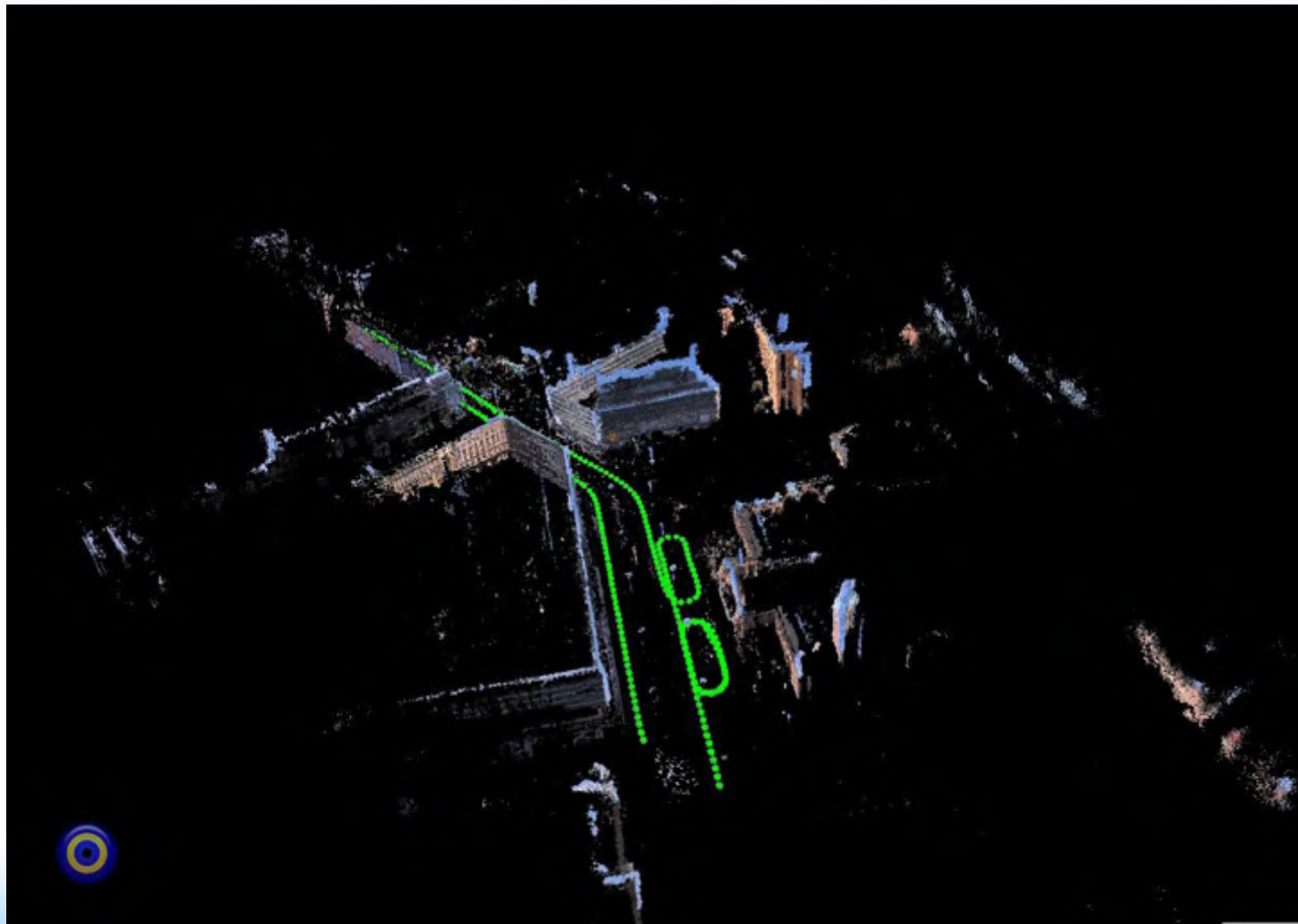
恒山悬空寺：位于山西浑源县，始建于公元491年，古代工匠根据道家“不闻鸡鸣犬吠之声”的要求建设了悬空寺，该寺距地面高约50米。





五台山佛光寺天地、室内外一体化三维重建

城市街景重建



城市重建



展望与趋势

- 三维视觉理论日趋成熟，逐渐走向应用
- 大数据、移动端
- 多传感器、几何与学习结合，提高鲁棒性与可靠性

谢谢

yhwu@nlpr.ia.ac.cn

<http://vision.ia.ac.cn/>