2017以来的2D to 3D

吴毅红，VALSE 2018，4月22日，大连

内容

- 前言
- 图像匹配
- 视觉定位， PnP，SLAM
- 三维重建，SfM，学习，RDGBD
- 发展趋势

前 言

- 3D 视觉应用：传统工业、AR、VR、机器人领域
  - 2016： AR、VR热潮
  - 2017：无人驾驶、机器人、AGV、3D摄像头
    - 2017.6.5，苹果 ARKit
    - 2017. 8，谷歌 ARCore
    - 2017.9.13， 苹果手机 iPhone X 3D 深度摄像头
    - 2017. 10. 19，三星和谷歌合作，ARCore引入三星手机；2018，华为、小米
    - 2017.11.28，华为手机荣耀 V10 配备3D 面部识别
    - 2018年3月，腾讯成立机器人实验室 Robotics X，张正友，三维视觉领域，三维重建、立体视觉、相机标定

图像匹配

- 概况
  - 传统设计的描述子逐渐被学习型描述子取代，深度学习成为主流趋势；
  - 深度学习开始在特征检测领域展现光彩；
  - 实际使用中，仍然是传统设计方法为主。

- 学习型特征检测算法
  - CovDet：CNN学习协变（Covariant）特征，Zhang, Yu, Kumar, Chang，CVPR2017
  - AffNet：CNN学习仿射协变参数，Mishkin，Radenovic，Matas，arXiv2017

学习型特征描述子

- L2Net：新的采样模式及误差，Tian, Fan, Wu，CVPR2017
- DeepCD：浮点描述子与二值描述子互补，Yang, Hsu, Lin, Chuang，ICCV2017
- Spread-out：学习描述子的空间分布，Zhang, Yu, Kumar, Chang，ICCV2017
- HardNet：基于L2Net的改进误差，Mishchuk, Mishkin, Radenovic, Matas，NIPS2017

特征匹配方法

- J. Bian, W. Lin, Y. Matsushita, S. Yeung, T. Nguyen, M. Cheng. GMS： Grid-based Motion Statistics for Fast, Ultra-robust Feature Correspondence. CVPR2017 (grid, 利用运动平滑信息进行特征匹配，快速、鲁棒)
- A. Seki, M. Pollefeys. SGM-Nets: Semi-global matching with neural networks. CVPR 2017. KITTI, Stereo 2012, rank 10

SGM: H. Hirschmuller. Stereo Processing by Semiglobal Matching and Mutual Information, PAMI 2008.

测评

- Johannes L. Schonberger, Hans Hardmeier, Torsten Sattler, Marc Pollefeys. Comparative Evaluation of Hand-Crafted and Learned Local Features. CVPR 2017.

- **新的数据库 Hpatches**

  描述子学习领域可用数据集较少，之前的Brown数据集性能趋于饱和

  HPatches在数据质量上较Brown数据集进一步提升，评测方法更有效、多样。

  HPatches A benchmark and evaluation of handcrafted and learned local descriptor. V. Balntas, K. Lenc, A. Vedaldi, K. Mikolajczyk. In CVPR 2017

视觉定位

- **3D点已知**

  2D to 3D 匹配，PnP，SLAM重定位

  T. Sattler, A. Torii, J. Sivic, M. Pollefeys, H. Taira, M. Okutomi, T. Pajdla, *Are Large-Scale 3D Models Really Necessary for Accurate Visual Localization?* CVPR 2017

- **3D点未知**

  SLAM，同步定位与地图构建

视觉定位全面的分类与综述

Yihong Wu, Fulin Tang, Heping Li. Image Based Camera Localization: an Overview. Invited Paper by Visual Computing for Industry, Biomedicine and Art, 2018.

3D点已知

大场景，异质数据

- Nathan Piasco, Désiré Sidibé, Cédric Demonceaux, Valérie Gouet-Brunet. A survey on Visual-Based Localization: On the benefit of heterogeneous data. Pattern Recognition, 2018.
- Liu Liu, Hongdong Li, and Yuchao Dai. Efficient Global 2D-3D Matching for Camera Localization in a Large-Scale 3D Map, ICCV 2017.
- Dylan Campbell, Lars Petersson, Laurent Kneip and Hongdong Li. Globally-Optimal Inlier Set Maximisation for Simultaneous Camera Pose and Feature Correspondence, ICCV 2017.
- Youji Feng, Yihong Wu, and Lixin Fan. Real-time SLAM Relocalization with On-line Learning of Binary Feature Indexing. Machine Vision and Applications, 2017.
- Jian Wu, Liwei Ma and Xiaolin Hu. Delving Deeper into Convolutional Neural Networks for Camera Relocalization. ICRA 2017.
- T. Qin, P. Li and S. Shen. Relocalization, Global Optimization and Map Merging for Monocualr Visual-Inertial SLAM. ICRA 2018.

3D点未知，SLAM

综述：

- C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J.J. Leonard. Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age. IEEE TRANSACTIONS ON ROBOTICS, 32(6), 2016.
- G. Younes, D. Asmar, E. Shammas, J. Zelek. Keyframe-based monocular SLAM: design, survey, and future directions. Robotics and Autonomous Systems 98 (2017) 67–88

复杂环境下的鲁棒视觉定位

通过点/线/面等多种几何元素的融合，解决弱纹理、强光线、长通道等环境下的视觉定位问题；通过多目/多传感器融合，提升复杂环境下的视觉定位精度和鲁棒性。

1. A. Pumarola, A. Vakhitov, et al. PL-SLAM: Real-Time Monocular Visual SLAM with Points and Lines. ICRA 2017.（点线）
2. S. C. Yang and S. Scherer. Direct Monocular Odometry Using Points and Lines. ICRA 2017. (点线)
3. S. C. Yang, Y. Song, et al. Pop-up SLAM: Semantic Monocular Plane SLAM for Low-texture Environments. IROS 2016.（平面）
4. P. F. Proenca and Y. Gao. Probabilistic rgb-d odometry based on points lines and planes under depth uncertainty.arXiv.org, 2017.（线面）
5. K. Sun, K. Mohta, et al. Robust Stereo Visual Inertial Odometry for Fast Autonomous Flight. IEEE Robotics and Automation Letters, 2018. (双目+IMU)
6. R. Wang, M. Schworer and D. Cremers. Stereo DSO: Large-Scale Direct Sparse Visual Odometry with Stereo Cameras. ICCV 2017. （双目）
7. K. Qiu, T. Liu and S. Shen. Model-based global localization for aerial robots using edge alignment. *IEEE Robotics and Automation Letters*, 2(3):1256-1263, 2017. （边缘）
8. Y. Ling, M. Kuse and S. Shen. Edge alignment-based visual-inertial fusion for tracking of aggressive motions. *Autonomous Robots*, pages 1-16, 2017. （边缘）
9. Y. Ling and S. Shen. Building maps for autonomous navigation using sparse visual SLAM features. IROS 2017.

学习或与几何融合的视觉定位

通过学习或与几何融合的方法提升视觉定位鲁棒性或精度：利用深度学习来估计深度和相机姿态，弥补几何方法在弱纹理区域的不足；同时，利用几何方法进一步提升深度和相机姿态的精度。

1. K. Tateno, F. Tombari, et al. CNN-SLAM: Real-time Dense Monocular SLAM with Learned Depth Prediction. CVPR 2017.
2. B. Ummenhofer, H. Z. Zhou, et al. DeMoN: Depth and Motion Network for Learning Monocular Stereo. CVPR 2017.
3. T. H. Zhou, M. Brown, N. Snavely, D.G. Lowe. Unsupervised Learning of Depth and Ego-Motion from video. CVPR 2017.
4. S. Vijayanarasimhan, S. Ricco, et al. Sfm-Net: Learning of Structure and Motion from Video. arXiv:1704.07804, 2017.
5. R. Li, S. Wang, et al. UnDeepVO: Monocular Visual Odometry through Unsupervised Deep Learning. arXiv: 1709.06841, 2017.
6. D. Detone, T. Malisiewicz, A. Rabinovich. Toward Geometric Deep SLAM. arXiv:1707.07410, 2017.
7. R. Clark, S. Wang, et al. VINet: Visual-Inertial Odometry as a Sequence-to-Sequence Learning Problem. AAAI 2017.
8. Xiang Gao, Tao Zhang. Unsupervised learning to detect loops using deep neural networks for visual SLAM system. Auton Robot (2017) 41:1–18.
9. Helder J. Araujo et al. Deep EndoVO: A Recurrent Convolutional Neural Network (RCNN) based Visual Odometry Approach for Endoscopic Capsule Robots, Neurocomputing, 2017.

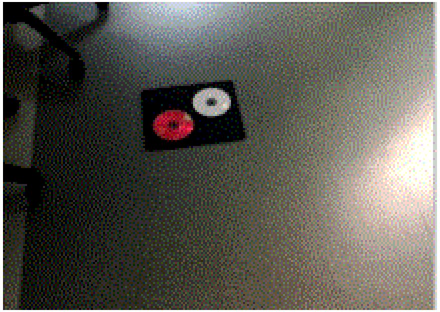
语义SLAM

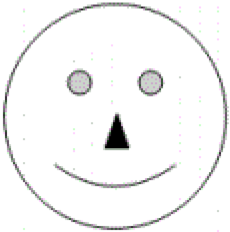
语义SLAM：从几何和内容两个层次感知世界，对地图内容进行抽象理解。通过语义理解，辅助SLAM提高建图和定位的精度；通过SLAM，帮助扩展语义理解的场景。

1. S. L. Bowman, N. Atanasov, et al. Probabilistic Data Association for Semantic SLAM. ICRA 2017. (Best Paper,5篇最佳论文之一)
2. J. McCormac, A. Handa, et al. Semantic Fusion: Dense 3D Semantic Mapping with Convolutional Neural Networks. ICRA 2017.

基于Marker的SLAM

1. Joseph DeGol, Timothy Bretl and Derek Hoiem. ChromaTag : A Colored Marker and Fast Detection Algorithm. ICCV, 2017.
2. R. Munoz-Salinas, M.J. Marin-Jimenez, E. Yeguas-Bolivar, R. Medina-Carnicer. Mapping and localization from planar markers. Pattern Recognition, Vol. 73, pp. 158-171, 2018.
3. Y. Wu. 轻量级无痕 marker SLAM. Patent, 2017. (不需要匹配，不需要PnP)





Event Camera SLAM， RGBD SLAM

基于事件相机的SLAM：事件相机的每个像素都在独立异步地感知接收光强变化，利用这种相机的低功耗、低带宽和对亮度变化非常敏感的性质来构建视觉SLAM系统。

1. G. Gallego, Jon E. A. Lund, et al. Event-based, 6-DOF Camera Tracking from Photometric Depth Maps. PAMI, 2017.
2. T. Rosinol Vidal, H. Rebecq, et al. Ultimate SLAM? Combining Events, Images, and IMU for Robust Visual SLAM in HDR and High Speed Scenarios. IEEE Robotics and Automation Letters, 2018
3. H. Rebecq, T. Horstschaefer and D. Scaramuzza. Real-time Visual-Inertial Odometry for Event Cameras using Keyframe-based Nonlinear Optimization. BMVC 2017.

Kinect，2009年6月，2012年发售，微软； 2017年，停产

- Yousif, K.; Taguchi, Y.; Ramalingam, S. MonoRGBD-SLAM: Simultaneous Localization and Mapping Using Both Monocular and RGBD Cameras. ICRA 2017.
- Helder J. Araujo et al. A Non-Rigid Map Fusion-Based RGB-Depth SLAM Method for Endoscopic Capsule Robots, 2017.

深度学习方法呈上升趋势
传统几何方法热情不减
实际应用还是传统的多视几何方法为主导

三维重建

从运动回复结构：SFM

学习

- 单目：直接学习深度
- 多目：学习匹配

深度相机：有限的可视范围和视角，大范围或者完整模型RGBD SLAM,非刚体

从运动恢复结构（SFM）

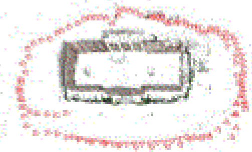
特征点检测与匹配

外极几何图构造

摄像机位姿和场景结构估计

- 增量式
- 全局式
- 混合式

捆绑调整 优化摄像机位姿和场景



点云处理

Onur Ozyesil, Vladislav Voroninski, Ronen Basri, Amit Singer
Survey on Structure from Motion
Acta Numerica， 2017

增量式进展

1. Hainan Cui, Shuhan Shen, Xiang Gao, Zhanyi Hu. Batched Incremental Structure-from-Motion. 3DV 2017
2. Jianwei Li, Wei Gao, Yihong Wu. Elaborate Scene Reconstruction with a Consumer Depth Camera. International Journal of Automation and Computing, 2017.

全局式进展

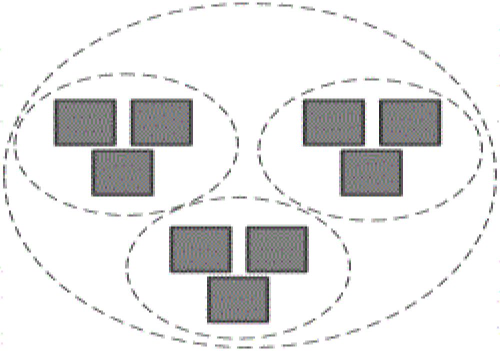
1. Hainan Cui, Shuhan Shen, Zhanyi Hu. Global Fusion of Generalized Camera Model for Efficient Large-Scale Structure from Motion. Science China: Information Sciences, 60: 038101:1–038101:3, 2017.

混合式进展

第一种：将摄像机位置和姿态求取拆开

Hainan Cui, Xiang Gao, Shuhan Shen, Zhanyi Hu. HSfM: Hybrid Structure-from-Motion. CVPR 2017.

第二种：将摄像机进行分组，每一组进行增量式重建，组之间再进行模型对齐时采用全局式方法。



1. Hainan Cui, Shuhan Shen, Xiang Gao, Zhanyi Hu. CSfM: Community-based Structure from Motion， ICIP 2017.
2. Siyu Zhu, Tianwei Shen, Lei Zhou, Runze Zhang, Jinglu Wang, Tian Fang, Long Quan. Parallel Structure from Motion from Local Increment to Global Averaging. ICCV 2017.

捆绑调整

1. Runze Zhang, Siyu Zhu, Tian Fang, Long Quan. Distributed Very Large Scale Bundle Adjustment by Global Camera Consensus. 29-38, ICCV 2017.
2. Hainan Cui, Shuhan Shen, Zhanyi Hu. Tracks Selection for Robust, Efficient and Scalable Large-Scale Structure from Motion. PR 2017.

天地融合

1. Lei Zhou, Siyu Zhu, Tianwei Shen, Jinglu Wang, Tian Fang, Long Quan. Progressive Large Scale-Invariant Image Matching in Scale Space. ICCV 2017.
2. Xiang Gao, Lihua Hu, Hainan Cui, Shuhan Shen, Zhanyi Hu. Accurate and Efficient Ground-to-Aerial Model Alignment. Pattern Recognition, 76(4): 288-302, 2018.
3. Yang Zhou, Shuhan Shen, Xiang Gao, Zhanyi Hu. Accurate Mesh-based Alignment for Ground and Aerial Multi-view Stereo Models. ICIP 2017.

点云处理

1. Nan. et al., PolyFit: Polygonal Surface Reconstruction from Point Clouds, ICCV 2017
2. Kelly. et al., BigSUR: Large-scale Structured Urban Reconstruction, TOG 2017
3. Zhu. et al., Variational Building Modeling from Urban MVS Meshes, 3DV 2017

学习深度

1. Clement Godard, Oisin Mac Aodha, Gabriel J. Brostow. Unsupervised Monocular Depth Estimation With Left-Right Consistency. CVPR 2017. (左右视图视差的一致性)

2. T. H. Zhou, M. Brown, N. Snavely, D.G. Lowe. Unsupervised Learning of Depth and Ego-Motion from video. CVPR 2017. (同时估计当前帧的深度以及相邻帧的相机相对姿态, 前边SLAM提过)

3. Lei He, Guanghui Wang and Zhanyi Hu. Learning Depth from Single Images with Deep Neural Network Embedding Focal Length, IEEE Transactions on Image Processing, 2018. 将固定焦距数据集转换成了多焦距数据集。焦距对深度估计有歧义性，消除由焦距信息引起的歧义性考虑到这种歧义性是由全局信息决定的，本文将焦距信息以全连接层的形式嵌入到全局特征网络中。

KITTI，双目

非刚体：

1. Kumar, Suryansh and Dai, Yuchao and Li, Hongdong. The 1st Winner of "Non-Rigid Structure from Motion Challenge 2017" @ CVPR 2017
2. Kumar, Suryansh and Dai, Yuchao and Li, Hongdong. Spatial-temporal union of subspaces for multi-body non-rigid structure-from-motion. Pattern Recognition, 2017. (An unified framework to jointly segment and reconstruct multiple non-rigid objects, along both temporal direction and spatial direction)
3. Suryansh, Hongdong Li et. al. Monocular Dense 3D Reconstruction of a Complex Dynamic Scene from Two Perspective Frames, ICCV 2017.
4. Kangkan Wang, Guofeng Zhang, Shihong Xia. Templateless Non-Rigid Reconstruction and Motion Tracking With a Single RGB-D Camera. IEEE Transactions on Image Processing, 26(12): 5966 – 5979, 2017.

其余：

1. T. Schoeps, T. Sattler, C. Haene, M. Pollefeys. Large-scale outdoor 3D reconstruction on a mobile device. CVIU 2017. （filter based method）
2. C. Haene, C. Zach, A. Cohen, M. Pollefeys, Dense Semantic 3D Reconstruction, PAMI, 2017.（ volumetric ）
3. Zhaopeng Cui, Jinwei Gu, Boxin Shi, Ping Tan and Jan Kautz. Polarimetric Multi-View Stereo. CVPR 2017. （从偏振，光度，法向信息，无纹理进行重建）

发展趋势

- **几何与学习融合**：传统多视几何主导三维视觉，辅以深度学习
- **多传感器融合**：视觉主导，辅以廉价激光、IMU等
- **与硬件结合**：深度相机，3D摄像头

  嵌入式SLAM时代来临？

  M. Abouzahir, A. Elouardi, R. Latif, S. Bouaziz, A. Tajer. Embedding SLAM algorithms: Has it come of age? Robotics and Autonomous Systems 100 (2018) 14–26.
- **与具体应用结合**：

  AGV，无人驾驶，服务机器人，AR教育，AR影音

谢 谢

yhwu@nlpr.ia.ac.cn
http://vision.ia.ac.cn/