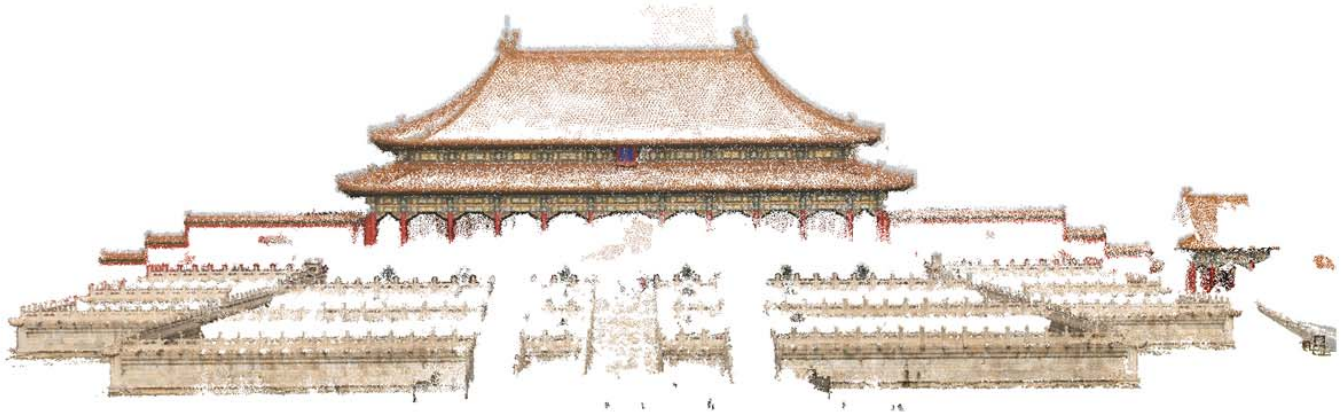
2016 Visual SLAM Report

Ping Tan

Simon Fraser University,
Qihu 360 AI Institute

Simultaneous Localization and Mapping (SLAM)



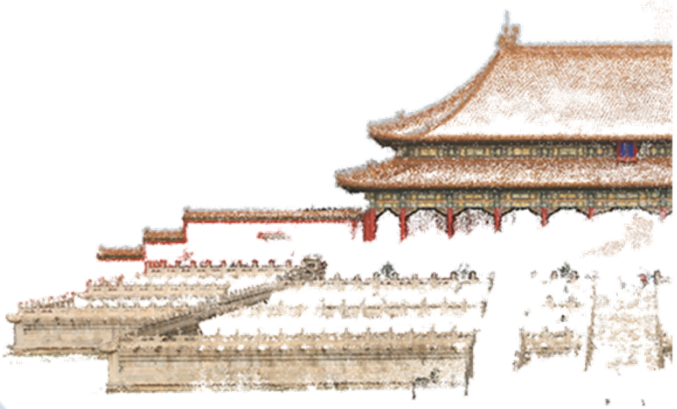
SLAM Components

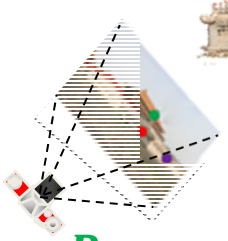
- Mapping (local)
  - Map reconstruction
- Tracking
  - Map-to-image registration
- Loops (global mapping)
  - Loop detection
  - Loop optimization

Odometry

SLAM Categorization

- By tracking objective
  - Feature-based method
    - ORB-SLAM, PTAM, etc.
  - Direct method
    - LSD-SLAM, DSO
  - Semi-direct method
    - SVO
- By sensor type
  - Camera (VO)
    - ORB-SLAM, LSD-SLAM, etc.
    - Event-camera SLAM (ECCV 2016 best paper)
  - Camera + IMU (VIO)
    - OKVIS, ROVIO, VI-ORB, RKSLAM
  - Depth camera (Fusion)
    - KinectFusion, BundleFusion, etc.





$\boldsymbol{R}, \mathbf{c}$

$$\rho(|\boldsymbol{Project}(\boldsymbol{Map}, \boldsymbol{R}, \mathbf{c}) - \boldsymbol{Observ}|)$$

Major Progress in 2016

- By the tracking objective
  - ORB-SLAM2 (feature based method)
  - DSO (direct method)
  - SVO2 (semi-direct method)
- By the sensor
  - ORB-SLAM2, DSO, SVO2 (Camera)
  - VI-ORB, RKSLAM (Camera + IMU)
  - BundleFusion (Depth Camera)

ORB-SLAM2

What's new from ORB-SLAM:

Support stereo and depth cameras

– New tracking objective function for stereo

$$\rho\left(\left|x^i_{(\cdot)} - \pi_{(\cdot)}\left(RX^i + t\right)\right|\right)$$

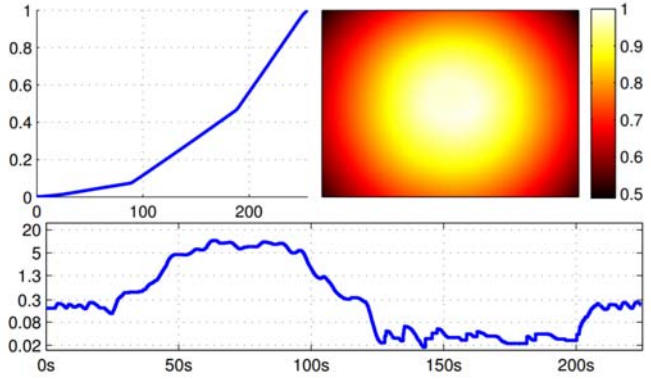
$$x_s = (u_L, v_L, u_R)$$

$$\pi_s\left(\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}\right) = \left[f_x \frac{X}{Z} + c_x, f_y \frac{Y}{Z} + c_y, f_x \frac{X - b}{Z} + c_x\right]$$

– A new formulation of depth camera tracking

- Synthesize u_R according to (u_L, v_L) and depth
- Adopt the stereo slam formulation

DSO



What's new from LSD-SLAM

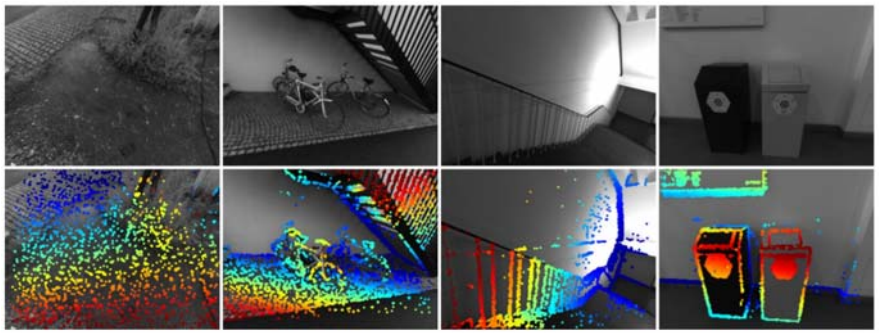
– Auto photometric calibration on the fly

$$\rho\left(\left(I_j[\boldsymbol{p}'] - b_j\right) - \frac{t_j e^{a_j}}{t_i e^{a_i}}\left(I_i[\boldsymbol{p}] - b_i\right)\right)$$

– Photometric BA (solving both pts and cameras)

- LSD solves pts and cameras iteratively

– Point sampling



Direct Sparse Odometry, J. Engel, V. Koltun, D. Cremers, arXiv Oct 2016

SVO2

What's new from SVO

– Support multi-camera rig, wide FoV lens

$$\rho\left(I_k^{\boxed{c}}\left(\pi\left(\boxed{T_{CB}}T_{k,k-1}\boldsymbol{p}\right)\right)-I_{k-1}^{\boxed{c}}\left(\pi\left(\boxed{T_{CB}}\boldsymbol{p}\right)\right)\right)$$

– Include motion priors

$$T_{k,k-1}=\arg\min\sum_{c}\sum_{p\in R_{k-1}}\rho(\cdots)^2+\boxed{\left|T_{k,k-1}-\tilde{T}_{k,k-1}\right|^2}$$

– Use of edgelet features

SVO: Semi-Direct Visual Odometry for Monocular and Multi-Camera Systems, C. Forster, Z. Zhang, M. Gassner, M. Werlberger, D. Scaramuzza, IEEE Transactions on Robotics, 2016

Better SLAM by Hardware or Software?

- DSO is more robust to motion blur
  - requires good lens, global shutter
- SVO is computationally efficient
  - works better on high fps camera
- VIO is more robust to blur & quick motion
  - requires good device synchronization
- Stereo SLAM is free from scale drift
  - requires two synchronized cameras
- Event camera is promising for brightness changes & quick motion
  - requires special hardware

Thank you!