Robust Monocular SLAM in Dynamic Environments

Wei Tan, Haomin Liu, Zilong Dong, Guofeng Zhang*, Hujun Bao

State Key Lab of CAD&CG, Zhejiang University
Roadmap

- Background and Related Work
- Key Issues for SLAM in Dynamic Scenes
- System Overview
- Online 3D Points and Keyframes Updating
- Prior-based Adaptive RANSAC
- Results and Comparison
Roadmap

- Background and Related Work
- Key Issues for SLAM in Dynamic Scenes
- System Overview
- Online 3D Points and Keyframes Updating
- Prior-based Adaptive RANSAC
- Results and Comparison
SLAM

- Simultaneous Localization and Mapping
  - Estimate the environment structure and the camera trajectory online, under a highly nonlinear partial observation model.
SLAM for Visual Odometry

SLAM for Augmented Reality
Multi-View Geometry

- Structure-from-Motion
  - Automatically recover the camera parameters and 3D structure from multiple images or video sequences.

Snavely et al. 2006
Multi-View Geometry

\[ x_{ij} = \pi \left( P_i X_j \right) \]

Projection Function

\[ \pi(x, y, z) = \left( \frac{x}{z}, \frac{y}{z} \right) \]

\[ P_i = K_i [R_i | T_i] \]
Structure-from-Motion

**Pipeline**

- **Feature Tracking**
  
  - Obtain a set of feature tracks

  \[ \mathcal{X} = \{ x_i \mid i = 1, \ldots, m \} \]

- **Structure from Motion**
  
  - Solve the camera parameters and 3D points of tracks

  \[ x_{ij} = \pi(P_i X_j) \]

  \[ P_i = K_i [R_i \mid T_i] \]

  \[ E(P_1, \ldots, P_m, X_1, \ldots, X_n) = \sum_{i=1}^{m} \sum_{j=1}^{n} w_{ij} \| \pi(P_i X_j) - x_{ij} \|^2 \]

**SLAM: real-time SfM**
Bundle Adjustment

Definition

- Refining a visual reconstruction to produce jointly optimal 3D structure and viewing parameter (camera pose and/or calibration) estimates.

\[
\arg \min_{p, x_i} \sum_{k=1}^{m} \sum_{i=1}^{n} D(x_{ki}, p_k(X_i))^2
\]

Real-Time Structure-from-Motion

- With a set of 2D-3D correspondences, quickly compute the camera pose by solving

\[ P_i = \arg \min_{P_i} \sum_j \| \pi(P_i X_j) - x_{ij} \|^2 \]

- The computational complexity is low and can be performed in real time.

- Parallel Tracking and Mapping
  - Foreground thread: track features and compute the camera pose with the estimated 3D points
  - Background thread: BA for map refinement
Related Work

- **Filter-based SLAM**
  - Davison et al. 2007, Eade and Drummond 2006

- **Keyframe-based SLAM**

- **SLAM in Dynamic Environments**
  - Shimamura et al. 2011, Zou and Tan, 2013

- **3D Change Detection**
  - Pollard and Mundy 2007, Taneja et al. 2011
Filter vs Keyframe BA

(a) Markov Random Field  (b) Filter  (c) Keyframe BA

Filter-based SLAM

- EKF State
  \[ x = \{C, X_1, \ldots, X_N\} \]

- Complexity \( O(N^3) \) per frame

- Poor scalability
  - Thousands points

Keyframe-based SLAM


PTAM
Key Issues for SLAM in Dynamic Environments

- Gradually changing
Key Issues for SLAM in Dynamic Environments

- Gradually changing
- Object Occlusion
  - Viewpoint Change
Key Issues for SLAM in Dynamic Environments

- Gradually changing
- Object Occlusion
- Viewpoint Change
- Dynamic Objects
Key Issues for SLAM in Dynamic Environments

- Gradually changing
- Object Occlusion
  - Viewpoint Change
  - Dynamic Objects
- Very low inlier ratio
Our Framework
Roadmap

- Background and Related Work
- Key Issues for SLAM in Dynamic Scenes
- System Overview
- Online 3D Points and Keyframes Updating
- Prior-based Adaptive RANSAC
- Results and Comparison
Online 3D Points and Keyframes Updating

- Keyframe representation
- 3D Change detection
  - Select 5 closest keyframes for online image.
  - For each valid feature point \( x \) in each selected keyframe,
    - Compute its projection \( x' \) in current frame
    - If \( n_x^T \hat{n}_{x'} < \tau_n \), compute the appearance difference
      \[
      D_c(X) = \min_d \sum_{y \in W(x)} |I_y - I_{y'+d}|
      \]
Online 3D Points and Keyframes Updating

- Keyframe representation
- 3D Change detection
  - Select 5 closest keyframes for online image.
  - For each valid feature point \( x \) in each selected keyframe,
    - Compute its projection \( x' \) in current frame
    - If \( n_x^T \hat{n}_{x'} < \tau_n \), compute the appearance difference \( D_c(X) = \min_d \sum_{y \in W(x)} |I_y - I_{y'+d}| \)
    - If \( D_c(X) > \tau_c \), then find a set of feature points \( y \) close to \( x' \).

Since dynamic points cannot be triangulated, the occlusion caused by dynamic objects can be excluded here.
Online 3D Points and Keyframes Updating

- **Keyframe representation**
- **3D Change detection**
  - Select 5 closest keyframes for online image.
  - For each valid feature point $x$ in each selected keyframe,
    - Compute its projection $x'$ in current frame
    - If $n_x^T \hat{n}_{x'} < \tau_n$, compute the appearance difference $D_c(X) = \min_d \sum_{y \in W(x)} |I_y - I_{y'} + d|$
      - If $D_c(X) > \tau_c$, then find a set of feature points $y$ close to $x'$.
        - If $z_{Xy} \geq z_X$ or their depths are very close, set $V(X) = 0$.

Since dynamic points cannot be triangulated, the occlusion caused by dynamic objects can be excluded here.

The occlusions caused by static objects are also excluded.
Occlusion Handling

Occlusions by Dynamic Objects

3D points updating with occlusion handling

3D points updating without occlusion handling

red points: invalid 3D points

http://www.cad.zju.edu.cn/home/gfzhang/projects/SLAM/ismar2013-paper293.wmv
Roadmap

- Background and Related Work
- Key Issues for SLAM in Dynamic Scenes
- System Overview
- Online 3D Points and Keyframes Updating
- Prior-based Adaptive RANSAC
- Results and Comparison
Prior-based Adaptive RANSAC

- Sample generation
  - 10x10 bins
  - Prior probability $p_i = \varepsilon_i^* / \sum_j \varepsilon_j^*$

- Hypothesis evaluation

\[ s = \left( \sum_i \varepsilon_i \right) \frac{\pi \sqrt{\det(C)}}{A} \]

- Inliers number $N \approx \sum_i \varepsilon_i$

- Inliers distribution, i.e., distribution ellipse $C$
Prior-based Adaptive RANSAC

- Hypothesis evaluation

\[ s = \left( \sum_i \varepsilon_i \right) \frac{\pi \sqrt{\det(C)}}{A} \]

\[ \sum_i \varepsilon_i = 24.94 \]

\[ \sum_i \varepsilon_i = 21.77 \]

200 green points on the static background, 300 cyan points on the rigidly moving object, 500 red points are randomly moving.
Prior-based Adaptive RANSAC

- Hypothesis evaluation

\[ s = \left( \sum_i \varepsilon_i \right) \frac{\pi \sqrt{\text{det}(C)}}{A} \]

\[ \sum_i \varepsilon_i = 24.94 \]

\[ \sum_i \varepsilon_i = 21.77 \]

S1 = 8.31 > S2 = 1.98

200 green points on the static background, 300 cyan points on the rigidly moving object, 500 red points are randomly moving.
Results and Comparison

http://www.cad.zju.edu.cn/home/gfzhang/projects/SLAM/ismar2013-paper293.wmv
RDSLAM: Robust Dynamic SLAM

Description

RDSLAM is a real-time simultaneous localization and mapping system which can robustly work in dynamic environments. It is for non-commercial research and educational use ONLY. Not for reproduction, distribution or commercial use. If you use this executable for your academic publication, please acknowledge our work. This program is tested on Win7, but is still not guaranteed to be bug-free and work properly with all versions of Windows. You are welcome to report any suggestions or bugs. We will actively update the program. Please email Guofeng Zhang if you have any questions.

Release (RDSLAM 1.0 released on Dec. 11, 2013)

RDSLAM 1.0 is implemented based on the following paper:


ChangeLog

http://www.zjucvg.net/rdslam/rdslam.html
Thank You!
Questions?