Evaluating Pose Estimation Methods for Stereo Visual Odometry on Robots

IAS Practice Talk

Hatem Alismail, Brett Browning and M. Bernardine Dias
{halismai,brettb,mbdias}@cs.cmu.edu

August 31, 2010
Visual Odometry

Intro. & Motivation | Pose Estimation | Evaluation | Results | Discussion | Conclusions
Stereo-based Visual Odometry Pipeline

Feature extraction, sparse stereo and tracking

- Many low quality corners computed efficiently
- Few high quality features, but cpu-intensive

Pose estimation

- IN: feature correspondences (3D/3D, 3D/2D)
- OUT: estimate of camera motion, R and T
  - RANSAC framework for robustness

Non-linear refinement

- Typically, minimize the reprojection error
  - Sparse Bundle Adjustment (SBA) is an optimal choice if noise is assumed to be Gaussian
Stereo-based Visual Odometry Pipeline

Feature extraction, sparse stereo and tracking

- Many low quality corners computed efficiently
- Few high quality features, but cpu-intensive

Pose estimation

- Many techniques are possible, which is the best for Visual Odometry?

Non-linear refinement

- Typically, minimize the reprojection error
  *Sparse Bundle Adjustment* (SBA) is an optimal choice if noise is assumed to be Gaussian

Intro. & Motivation | Pose Estimation | Evaluation | Results | Discussion | Conclusions
Accurate pose estimation is crucial for good VO
Good pose estimation speeds up the convergence of non-linear refinement to the correct solution
A robust pose estimation algorithm:
- Able to handle small number of features
  - Washed-out/saturated image, little textures, ...
- Avoid degenerate solutions
- Efficient
Pose Estimation

Absolute Orientation (AO)

- Rigid motion between two 3D point sets

\[ R^*, T^* = \text{argmin}_{R,T} \sum_{i}^{N} ||X_i - (RY_i + T)||^2 \]

- Closed form solutions, using unit quaternions (Horn, 87) and SVD decomposition (Umeyama, 93)

Perspective-N-Points (PnP)

- Motion of a camera given 3D world points and their corresponding 2D projections on the image

- Nonlinear solutions (Haralick et. al., 94)
- Linear solutions (Wu & Hu, 06) (Intell, et. al. 03) (Quan & Lan, 99)
# AO vs. PnP

## Absolute Orientation
- **Pros**
  - Very easy to implement
  - Very efficient
  - Camera-model independent
- **Cons**
  - Least squares estimation assumes Gaussian noise
    - Stereo triangulation noise is not Gaussian
    - Weighted AO based on 3D points covariance is possible, but results are similar

## Perspective-N-Points
- **Pros**
  - More accurate and robust
    - Does not rely directly on 3D structure, instead perspective projection constraints
  - Large number of efficient algorithms
- **Cons**
  - Requires central perspective camera model
  - Implementation more involved
Evaluation Framework

• Features
  – Harris corners ~1000-1500 per frame with binning (Nister et. al. 04)
  – Sparse stereo: ZNCC window matching on the epipolar line
  – Tracking: ZNCC matching between frames in a square window

• Pose estimation
  – Different algorithms in a RANSAC framework to be evaluated

• Non-linear minimization
  – SBA minimization of the reprojection error (Lourakis & Argyros, 04)
Evaluation Criteria

• **Loop closure**
  – Distance between start and finish position in a loop dataset

• **SBA reprojection error**
  – Lower error indicates better initial pose estimates

• **Number of RANSAC iterations to find a solution**
  – Indicates amount of computation needed to find a solution

• **Percentage of RANSAC inliers**
  – Indirect measure of robustness
### Algorithms

<table>
<thead>
<tr>
<th>AO</th>
<th>Absolute Orientation (Umeyama, 93)</th>
</tr>
</thead>
<tbody>
<tr>
<td>o-P3P</td>
<td>Original Perspective-3-Points (Fischler &amp; Bolles, 06)</td>
</tr>
<tr>
<td>l-P3P</td>
<td>Linear Perspective-3-Points (Intell, et. al. 03)</td>
</tr>
<tr>
<td>l-P4P</td>
<td>Linear Perspective-4-Points (Intell, et. al. 03)</td>
</tr>
<tr>
<td>E-P5P</td>
<td>Non-linear Efficient-P5P (Moreno-Noguer, 07)</td>
</tr>
</tbody>
</table>

Algorithms are first evaluated on the indoors datasets. The best is then evaluated on the outdoors datasets.

*All algorithms have been tested on simulated data for correctness*
Indoor Data

- Modified LAGR robot
- Gyroscope + wheel odometry
- 3.8mm lens, 12cm baseline
- Wean Hall, Carnegie Mellon

<table>
<thead>
<tr>
<th></th>
<th>Loop closure m(% total dist)</th>
<th>Initial SBA error (px)</th>
<th>Final SBA error (px)</th>
<th>RANSAC iterations</th>
<th>% RANSAC inliers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-P3P</td>
<td>6.68 (1.50%)</td>
<td>16.45</td>
<td>13.41</td>
<td>29.10</td>
<td>0.58</td>
</tr>
<tr>
<td>o-P3P</td>
<td>6.29 (1.36%)</td>
<td>16.48</td>
<td>13.44</td>
<td>29.09</td>
<td>0.58</td>
</tr>
<tr>
<td>EP5P</td>
<td>7.87 (1.72%)</td>
<td>16.91</td>
<td>13.49</td>
<td>174.85</td>
<td>0.56</td>
</tr>
<tr>
<td>1-P4P</td>
<td>2.2x10³ (46.03%)</td>
<td>25.92</td>
<td>19.97</td>
<td>509.47</td>
<td>0.42</td>
</tr>
</tbody>
</table>

- AO fails consistently on the indoor datasets (wide FOV stereo)
- 1-P3P and o-P3P are the best with statistically insignificant differences
- 1-P4P performs badly
- EP5P has an acceptable performance, but not as good as P3P
Indoor Data

- **Intro. & Motivation**
- **Pose Estimation**
- **Evaluation**
- **Results**
- **Discussion**
- **Conclusions**
AO performance on a simulated stereo frame

5 degrees yaw + [0.5 0.5 0.5] meter translation
Outdoor Data

- Camera mounted in a car
- GPS, with WASS correction
- 6mm lenses, 12cm baseline
- Education city, Qatar

<table>
<thead>
<tr>
<th></th>
<th>Loop closure m(% total dist)</th>
<th>Initial SBA error (px)</th>
<th>Final SBA error (px)</th>
<th>RANSAC iterations</th>
<th>% RANSAC inliers</th>
</tr>
</thead>
<tbody>
<tr>
<td>P3P</td>
<td>6.1 (1.1%)</td>
<td>69.29</td>
<td>59.49</td>
<td>96.70</td>
<td>0.41</td>
</tr>
<tr>
<td>AO</td>
<td>500 (8.1%)</td>
<td>88.74</td>
<td>61.73</td>
<td>$2.55 \times 10^3$</td>
<td>0.30</td>
</tr>
</tbody>
</table>

- AO’s performance improves (narrow FOV)
- P3P’s performance is still better
What about using more points?

- More points resolve ambiguity in the case of PnP
- RANSAC performance improves with a lower DOF model
Conclusions

• Selection of the right focal length/baseline is very important
• PnP is better than AO for pose estimation from stereo data
• P3P’s performance outperforms the rest of the algorithms

Acknowledgements

- This work was funded by Qatar foundation seed grant
- Thanks to Balajee Kannan, Freddie Dias, Jimmy Bourne, Victor Marmol and Dominic Jonak for their help with hardware and data collection
Questions?