Photometric Bundle Adjustment for Globally Consistent Mapping

Simon Klenk
Technische Universität München
Chair of Computer Vision & AI
Master Thesis
Advisor: M.Sc. Nikolaus Demmel
Supervisor: Prof. Dr. Daniel Cremers
Motivation: Improving Photometric Maps

Before Loop Closure

After Loop Closure

Direct Sparse Odometry with Loop Closure [1]

Stairs converges to one object

Even more Improvement:

Photometric Bundle Adjustment

Research Question:
Is the current implementation without alternative?

Evaluation:
Kitti odometry 00-10
Euroc MAV
PBA Cost Formulation: Direct Image Error

\[ p' = \pi \left[ T_{ji} \pi^{-1} (p, id_p) \right] \]
PBA Cost Formulation: Direct Image Error

\[ E_{\text{photo}} = \sum_{\text{frames}} \sum_{\text{points}} \sum_{\text{obs}} \sum_{\text{pattern}} \| I_j [p'] - I_i [p] \|_{Huber} \]

Point \( p \) in Host Frame \( i \)

Target Frame \( j \)

\[ p' = \pi [ T_{ji} \pi^{-1} (p, id_p) ] \]
Residual Pattern Geometry

Spherical Patterns (inverse distance)  
0.675 $ATE_{avg}$

Planar Patterns (inverse depth)  
0.684 $ATE_{avg}$

Which is better?
Residual Pattern: Normal Vectors

Initialization

How to optimize the normal vectors?

After normal vector optimization
Residual Pattern: Normal Vectors

<table>
<thead>
<tr>
<th></th>
<th>[PBA, normals]</th>
<th>[normals, PBA]</th>
<th>[PBA + normals]</th>
</tr>
</thead>
<tbody>
<tr>
<td>all sequences</td>
<td>0.672</td>
<td>0.762</td>
<td>0.731</td>
</tr>
</tbody>
</table>

enlarged

ground truth
init_pgo
[PBA+normals]
[PBA, normals]
[normals, PBA]
eurocV202
Where else did we have a closer look?

$$
E_{\text{photo}} = \sum_{\text{frames}} \sum_{\text{points}} \sum_{\text{obs}} \sum_{\text{pattern}} \| I_j [p'] - I_i [p] \|_{\text{Huber}}
$$
Host-Target Transformation: Interpolation in Target

Computing exact gradients

Computing smooth gradients: using gradient image (central differences)

Bilinear interpolation [2]
Host-Target Transformation: Interpolation in Target

![Graph showing Host-Target Transformation: Interpolation in Target](image)

<table>
<thead>
<tr>
<th>Method</th>
<th>ATE\textsubscript{rmse,geo}</th>
</tr>
</thead>
<tbody>
<tr>
<td>init_pgo</td>
<td>1.0</td>
</tr>
<tr>
<td>bilin</td>
<td>0.671</td>
</tr>
<tr>
<td>bilin_s</td>
<td>0.673</td>
</tr>
<tr>
<td>bicubic</td>
<td>0.669</td>
</tr>
<tr>
<td>bicubic_s</td>
<td>0.666</td>
</tr>
<tr>
<td>bicubic_smooth@20lt</td>
<td>0.657</td>
</tr>
</tbody>
</table>
Host-Target Transformation: Interpolation in Target

Smooth gradients are similar to interpolating on image pyramid
Where else did we have a closer look?

\[ T_{ji} = T_j T_i^{-1} \]

\[ \pi^{-1}(p, id_p) \]

\[ p' = \pi \left[ T_{ji} \pi^{-1}(p, id_p) \right] \]

\[ E_{photo} = \sum \sum \sum \sum \sum \| I_j[p'] - I_i[p] \|_{Huber} \]
Host-Target Transformation: Approximation

full warp:

\[ p'_k = \pi [ T_{ji} \pi^{-1} (p_0 + u_k, id_p) ] \]
Host-Target Transformation: Approximation

**full warp:**

\[ p_k' = \pi [T_{ji} \pi^{-1} (p_0 + u_k, id_p)] \]

**simple warp:**

\[ p_k' = \pi [T_{ji} \pi^{-1} (p_0, id_p)] + u_k \]
Host-Target Transformation: Approximation

**full warp:**

\[ p_k' = \pi \left[ T_{ji} \pi^{-1} (p_0 + u_k, id_p) \right] \]

**simple warp:**

\[ p_k' = \pi \left[ T_{ji} \pi^{-1} (p_0, id_p) \right] + u_k \]

**exact full warp:**
1) Warp all exactly

**approximate full warp:**
1) Warp by 1\textsuperscript{st} order Taylor at \( p_0 \)
2) Jacobian only for central pixel \( p_0 \)
Host-Target Transformation: Approximation

**full warp:**

\[ p'_k = \pi [T_{ji} \pi^{-1} (p_0 + u_k, id_p)] \]

**exact ↔ approximate**

**simple warp:**

\[ p'_k = \pi [T_{ji} \pi^{-1} (p_0, id_p)] + u_k \]

<table>
<thead>
<tr>
<th>warp:</th>
<th>DSO pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>exact</td>
</tr>
<tr>
<td>all</td>
<td>0.707</td>
</tr>
</tbody>
</table>
Host-Target Transformation: Approximation

\[
p_k' = \pi [T_{ji} \pi^{-1} (p_0 + u_k, id_p)]
\]

**full warp:**

<table>
<thead>
<tr>
<th>warp:</th>
<th>DSO pattern</th>
<th>exact</th>
<th>approx</th>
<th>simple</th>
</tr>
</thead>
<tbody>
<tr>
<td>all</td>
<td>0.707</td>
<td>0.693</td>
<td>0.693</td>
<td></td>
</tr>
<tr>
<td>euroc-ok</td>
<td>0.688</td>
<td>0.691</td>
<td>0.690</td>
<td></td>
</tr>
<tr>
<td>euroc-fail</td>
<td>1.183</td>
<td>0.997</td>
<td>1.006</td>
<td></td>
</tr>
</tbody>
</table>

**simple warp:**

\[
p_k' = \pi [T_{ji} \pi^{-1} (p_0, id_p)] + u_k
\]
Host-Target Transformation: Approximation

**full warp:**
\[ p_k' = \pi [T_{ji} \pi^{-1} (p_0 + u_k, id_p)] \]

**simple warp:**
\[ p_k' = \pi [T_{ji} \pi^{-1} (p_0, id_p)] + u_k \]

<table>
<thead>
<tr>
<th>warp:</th>
<th>DSO pattern</th>
<th>exact</th>
<th>approx</th>
<th>simple</th>
</tr>
</thead>
<tbody>
<tr>
<td>all</td>
<td></td>
<td>0.707</td>
<td>0.693</td>
<td>0.693</td>
</tr>
<tr>
<td>euroc-ok</td>
<td></td>
<td>0.688</td>
<td>0.691</td>
<td>0.690</td>
</tr>
<tr>
<td>euroc-fail</td>
<td></td>
<td>1.183</td>
<td>0.997</td>
<td>1.006</td>
</tr>
<tr>
<td>kit-no-loop</td>
<td></td>
<td>0.719</td>
<td>0.724</td>
<td>0.702</td>
</tr>
<tr>
<td>kit-loop</td>
<td></td>
<td>0.548</td>
<td>0.568</td>
<td>0.579</td>
</tr>
</tbody>
</table>
Host-Target Transformation: Approximation

\[ p_k' = \pi [T_{ji} \pi^{-1} (p_0 + u_k, id_p)] \]

**full warp:**

\[ p_k' = \pi [T_{ji} \pi^{-1} (p_0 + u_k, id_p)] \]

**exact** ↔ **approximate**

**simple warp:**

\[ p_k' = \pi [T_{ji} \pi^{-1} (p_0, id_p)] + u_k \]

<table>
<thead>
<tr>
<th>warp</th>
<th>DSO pattern</th>
<th>9x9 sparse</th>
<th>13x13 sparse</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>exact</td>
<td>approx</td>
<td>simple</td>
</tr>
<tr>
<td>all</td>
<td>0.707</td>
<td>0.693</td>
<td>0.693</td>
</tr>
<tr>
<td>euroc-ok</td>
<td>0.688</td>
<td>0.691</td>
<td>0.690</td>
</tr>
<tr>
<td>euroc-fail</td>
<td>1.183</td>
<td>0.997</td>
<td>1.006</td>
</tr>
<tr>
<td>kit-no-loop</td>
<td>0.719</td>
<td>0.724</td>
<td>0.702</td>
</tr>
<tr>
<td>kit-loop</td>
<td>0.548</td>
<td>0.568</td>
<td>0.579</td>
</tr>
</tbody>
</table>
Host-Target Transformation: Approximation

**full warp:**
\[ p_k' = \pi [T_{ji} \pi^{-1} (p_0 + u_k, id_p)] \]

**simple warp:**
\[ p_k' = \pi [T_{ji} \pi^{-1} (p_0, id_p)] + u_k \]

<table>
<thead>
<tr>
<th>warp:</th>
<th>DSO pattern</th>
<th></th>
<th>9x9 sparse</th>
<th></th>
<th>13x13 sparse</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>exact</td>
<td>approx</td>
<td>simple</td>
<td>exact</td>
<td>approx</td>
<td>simple</td>
</tr>
<tr>
<td>all</td>
<td>0.707</td>
<td>0.693</td>
<td>0.693</td>
<td>0.739</td>
<td>0.743</td>
<td>0.787</td>
</tr>
<tr>
<td>euroc-ok</td>
<td>0.688</td>
<td>0.691</td>
<td>0.690</td>
<td>0.738</td>
<td>0.731</td>
<td>0.779</td>
</tr>
<tr>
<td>euroc-fail</td>
<td>1.183</td>
<td>0.997</td>
<td>1.006</td>
<td>0.996</td>
<td>0.996</td>
<td>0.986</td>
</tr>
<tr>
<td>kit-no-loop</td>
<td>0.719</td>
<td>0.724</td>
<td>0.702</td>
<td>0.778</td>
<td>0.781</td>
<td>0.766</td>
</tr>
<tr>
<td>kit-loop</td>
<td>0.548</td>
<td>0.568</td>
<td>0.579</td>
<td>0.566</td>
<td>0.566</td>
<td>0.603</td>
</tr>
</tbody>
</table>
Where else did we have a closer look?

\[ E_{\text{photo}} = \sum_{\text{frames}} \sum_{\text{points}} \sum_{\text{obs}} \sum_{\text{pattern}} \| I_j [p'] - I_i [p] \|_t - \text{distribution} \]
Robust Norms: t-distribution

ours: \( W_{i,t} = \frac{1}{\sigma_t^2} \frac{v+1}{v+(\frac{r_i}{\sigma_t})^2} \)

old [3]: \( W_{i,t} = \frac{v+1}{v+(\frac{r_i}{\sigma_t})^2} \)

cost: \( c = \sum_i W_{i,t} r_i^2 \)
Robust Norms: t-distribution

\[ c = \sum_{i} w_i r_i^2 \]

<table>
<thead>
<tr>
<th>weight:</th>
<th>TDist weight</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>all sequences</td>
<td>( w_{i,corrected} )</td>
<td>( w_{i,old} )</td>
</tr>
<tr>
<td>euroc-ok</td>
<td>0.627</td>
<td>0.702</td>
</tr>
<tr>
<td>euroc-fail&amp;eurocV202</td>
<td>1.355</td>
<td>1.053</td>
</tr>
<tr>
<td>kit-no-loop</td>
<td>0.520</td>
<td>0.584</td>
</tr>
<tr>
<td>kit-loop</td>
<td>0.720</td>
<td>0.683</td>
</tr>
</tbody>
</table>

ours: \( W_{i,t} = \frac{1}{\sigma_t^2} \frac{v+1}{v+\left(\frac{r_i}{\sigma_t}\right)^2} \)

old [3]: \( W_{i,t} = \frac{v+1}{v+\left(\frac{r_i}{\sigma_t}\right)^2} \)
Robust Norms: t-distribution

\[ c = \sum_i w_{i,t} r_i^2 \]

<table>
<thead>
<tr>
<th>weight:</th>
<th>( w_{i,\text{corrected}} )</th>
<th>( w_{i,\text{old}} )</th>
<th>average CoV</th>
</tr>
</thead>
<tbody>
<tr>
<td>all sequences</td>
<td>0.708</td>
<td>0.708</td>
<td>0.58</td>
</tr>
<tr>
<td>euroc-ok</td>
<td>0.627</td>
<td>0.702</td>
<td>0.65</td>
</tr>
<tr>
<td>euroc-fail&amp;eurocV202</td>
<td>1.355</td>
<td>1.053</td>
<td>0.74</td>
</tr>
<tr>
<td>kit-no-loop</td>
<td>0.520</td>
<td>0.584</td>
<td>0.51</td>
</tr>
<tr>
<td>kit-loop</td>
<td>0.720</td>
<td>0.683</td>
<td>0.41</td>
</tr>
</tbody>
</table>

\[
\text{CoV} = \frac{\text{var}([\sigma])}{\text{mean}([\sigma])}
\]

ours: \[
W_{i,t} = \frac{1}{\sigma_t^2} \frac{v+1}{v+\left(\frac{r_i}{\sigma_t}\right)^2}
\]

old [3]: \[
W_{i,t} = \frac{v+1}{v+\left(\frac{r_i}{\sigma_t}\right)^2}
\]
Where else did we have a closer look?

\[ E_{\text{photo}} = \sum_{\text{frames}} \sum_{\text{points}} \sum_{\text{obs}} \sum_{\text{pattern}} \left\| I_j[p'] - I_i[p] \right\|_{\text{Huber}} \]
Residual Formulations

- Explicit brightness model (per image): \( ABOPT \)

\[
\mathbf{r}_{ab}^{(k)} = (I_j[p'_k] - b_j) - \frac{e^{a_j}}{e^{a_i}} (I_i[p_k] - b_i)
\]

- Implicit brightness model (per patch): \( LSSD, LNSSD, ZNCC/ZNSSD \)

\[
\mathbf{r}_{lssd}^{(k)} = I_j[p'_k] - \frac{\overline{I}_j}{\overline{I}_i} I_i[p_k]
\]

<table>
<thead>
<tr>
<th>residuals:</th>
<th>SSD</th>
<th>LSSD</th>
<th>LNSSD</th>
<th>ABOPT</th>
</tr>
</thead>
<tbody>
<tr>
<td>all sequences</td>
<td>0.693</td>
<td>0.658</td>
<td>0.670</td>
<td>0.666</td>
</tr>
</tbody>
</table>
Residual Formulations

- Explicit brightness model (per image): $ABOPT$

$$r_{ab}^{(k)} = (I_j[p'_k] - b_j) - \frac{e^{a_j}}{e^{a_i}}(I_i[p_k] - b_i)$$

- Implicit brightness model (per patch): $LSSD, LNSSD, ZNCC/ZNSSD$

$$r_{lssd}^{(k)} = I_j[p'_k] - \frac{\bar{I}_j}{\bar{I}_i} I_i[p_k]$$

$$2 \times (1 - ZNCC) = ZNSSD$$

<table>
<thead>
<tr>
<th>residuals:</th>
<th>SSD</th>
<th>LSSD</th>
<th>LNSSD</th>
<th>ABOPT</th>
<th>ZNCC</th>
<th>ZNSSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>all sequences</td>
<td>0.693</td>
<td>0.658</td>
<td>0.670</td>
<td>0.666</td>
<td>0.751</td>
<td>0.676</td>
</tr>
</tbody>
</table>
Overview of other experiments

- **Huber:**
  - Per-target frame works, with different scale estimator (same as for t-distribution, MAD, or sample standard deviation tested)

- **Self-tuning M-estimation [4]:**
  - Achieves very good results for t-distribution
  - Most general and therefore preferred

- **LM dampening:**
  - No big difference between options, most efficient should be used, e.g. only landmark dampening (identity or original Hessian or Schur)

- **LM step criteria:**
  - Okay to evaluate PBA cost or linearized costs, theoretically OLS correct

- **Triggs correction:**
  - Second order correction of Hessian for robust loss
  - Small improvement for t-distribution, for Huber not because only outlier contribute to corrected Hessian

- **Occlusion geometric & photometric:**
  - Simple approaches results only in very minor improvement
Conclusions

- **Use** residuals which account for brightness changes
- **Use** smooth gradients in the beginning, exact gradients in the end
- **Use** full warp: approximated version is usually fine, simple warp is too simple
- **Use** normal optimization as separate step after PBA
- **Use** self-tuning approach (or corrected formula for t-distribution)
- **Use** Triggs-correction for t-distribution case
- **Use** any kind of dampening (diagonal of Hessian/Schur or identity)

- **Future Work:**
  - different metrics required! (especially map evaluation)
  - Numerical properties
  - Occlusion detections / Deduplication
  - Benchmark on more data & against DL / feature-based
Thanks for listening and asking questions!
Sources

[1] X. Gao, R. Wang, N. Demmel and D. Cremers, LDSO: Direct Sparse Odometry with Loop Closure, iros, October 2018

