BA-NET: Dense Bundle Adjustment Networks.

Background on Bundle Adjustment

1. Given images $\{I_i\}_{i=1}^{N_i}$,
   Let the camera poses be $\{T_i\}_{i=1}^{N_i}$ and
   the 3D scene point coordinate be $\{p_j\}_{j=1}^{N_j}$

2. Geometric BA jointly optimizes for the poses and coordinates
   by minimizing reprojection error:
   $$X = \text{argmin} \sum_i \sum_j \| e^g_{i,j}(X) \|$$

3. $e^g_{i,j}(X) = \pi(T_i, p_j) - q_{i,j}$
   measures the diff. b/t a projected scene point & its corresponding feature point.

4. π projects scene points to image space

5. $q_{i,j} = [x_{i,j}, y_{i,j}, 1]$ is the normalized homogeneous pixel coordinate.

6. $X = [T_1, T_2, \dots, T_{N_i}, p_1, p_2, \dots p_{N_j}]^T$ ← camera poses & points coordinates

Scanned with CamScanner

7. Another approach is photometric BA which directly minimizes photometric error (pixel intensity difference) of aligned pixels.

8. The photometric error is:

$$e^{p}_{i,j}(\chi) = I_i(\pi(T_i, d_j \cdot q_j)) - I_1(q_j)$$

9. d_j is the depth of pixel q_j at the image I_1.

10. $d_j \cdot q_j$ = 3D coordinate of pixel q_j

11. Thus, $\chi = [T_1, T_2, \ldots, T_{N_i}, d_1, d_2, \ldots, d_{N_j}]$

BA-Net architecture

1. Their work is similar to photometric BA, but they minimize the feature-metric difference of aligned pixels:

$$e^{f}_{i,j}(\chi) = F_i(\pi(T_i, d_j \cdot q_j)) - F_1(q_j)$$

2. where $\{F_i\}_{i=1}^{N_i}$ are the feature pyramid of images $\{I_i\}_{i=1}^{N_i}$

3. They learn F, instead of using pretrained features.

4. They therefore introduce the BA-Net layer to solve the argmin optimization problem, so that the loss can be back-propagated.

5. The BA-Layer predicts the camera poses $\{T_i\}$ and the dense depth map $\{d_j\}$ during forward pass.

6. During backward pass, the BA-Layer backprop loss from TE D to the feature pyramid F for training.

Scanned with CamScanner