Objective

Given:
1. Three 2D-3D lines correspondences $l_i \leftrightarrow L_W$, $i = 1, 2, 3$.
2. $l_i$ are seen respectively by three cameras $F_{C_i}$, $i = 1, 2, 3$ and $L_W$ is defined in a fixed world frame $F_W$.
3. $F_{C_i}$ are rigidly fixed together with known camera intrinsics $K_i$ and extrinsics $T_{F_{C_i}}^{F_W}$, $i = 1, 2, 3$.

Find: The pose of the multi-camera system with respect to the fixed world frame, i.e. relative transformation $(R_{W}^{C_1}, t_{W}^{C_1})$ that brings a point defined in the multi-camera reference frame $F_{C_1}$ to the fixed world frame $F_W$.

Plücker Line Representation

- 2D-3D line correspondence $l_C \leftrightarrow L_W$.
- $P_{W}^{C_1}$, $P_{W}^{C_2}$: end-points of 3D line $L_W$.
- $p_{a_i}, p_{b_i}$: end-points of 2D line $l_C$.

NP3L: 3-Line Minimal Solution

Solving for $R_{W}^{C_1}$

- Since $U^TV$ is zero for any Plücker line $[U^TV]^T$, $u_C$ is parallel to $U_C$, and $R_{W}^{C_1} = R_{C_1}R_{W}^{C_1}$, from $L_C \leftrightarrow l_C$ we get
  $$u_C^R_{W}R_{W}^{C_1}V_W = 0.$$  

  (2)

- Rewrite Eq. (2) into $Ar = 0$, $a$ is a $1 \times 9$ matrix from known variables $u_C^R$, $R_{C_1}$ and $V_W$, and $r$ is a 9-vector from the unknown rotation matrix $R_{W}^{C_1}$.

  Given three 2D-3D line correspondences in the minimal case, we get $Ar = 0$ where $A$ is a $3 \times 9$ matrix from $a_j/y_j = 1, 2, 3$. Right null-space of $A$ gives
  $$r = \beta_1b_1 + \beta_2b_2 + \beta_3b_3 + \beta_4b_4 + \beta_5b_5 + \beta_6b_6.$$  

  (3)

- Setting $\beta = 1$ and enforcing the orthogonality constraint on the rotation matrix formed by $r$, we get a system of 10 polynomial equations in terms of the unknowns $\beta_1, ... , \beta_6$.

  $$f_j(\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6) = 0, \quad j = 1, 2, ..., 10.$$  

  (4)

which can be solved with Gröbner basis to get eight solutions.

Solving for $t_{W}^{C_1}$

- Since $U_C$ and $u_C$ are parallel, from $L_C \leftrightarrow l_C$ we get
  $$\lambda u_C = (R_{C_1}R_{W}^{C_1} \quad R_{C_1}t_{C_1}^{W} + t_{W}^{C_1}R_{W}^{C_1}) (V_W^{T} \quad V_W).$$  

  (5)

- Taking cross-product of $u_C$ on both sides to get rid of the unknown scalar value $\lambda$ and with three 2D-3D line correspondences, we get $Bt = 0$.

- Solution of the unknown $t = (t_{W}^{C_1} \quad t_{W}^{C_2} \quad t_{W}^{C_3})^T$ is given by the right null-space of $B$ which is made of known variables $R_{C_1}, t_{C_1}^{W}, R_{C_2}, t_{C_2}^{W}, u_C, U_W$ and $V_W$.

NPnL: $\geq 3$ Line Correspondences

- For $n \geq 3$ 2D-3D line correspondences, $A$ and $B$ are $n \times 9$ and $2n \times 4$ matrices. The solution steps remain the same as the minimal case.

Special Cases

- One Camera: becomes the perspective pose estimation problem when all line correspondences are seen by only one camera. Camera extrinsics $(R_{C_2}, t_{C_2})$ vanishes and we directly solve for the camera pose $(R_{W}^{C_1}, t_{W}^{C_1})$ without any change to the algorithm.

- Parallel 3D Lines: minimal case where two or all the three lines are parallel. Rank of matrix $A$ drops below 3 and $R_{W}^{C_1}, t_{W}^{C_1}$ cannot be solved. Fortunately, we can prevent this degenerate case by omitting parallel lines.

Results

Simulations (Non-Perspective):

- $R_{C_1}^C, R_{C_2}^C$: end-points of 3D line transformed into camera frame.
- $p_{a_i}, p_{b_i}$: reprojection of $R_{C_1}^C, R_{C_2}^C$ onto the image.
- $P_{C_1}, P_{C_2}$: camera matrix normalized 2D line end-points.

Reprojection Error for RANSAC

$$e = \frac{d_a^2 + d_b^2}{2||p_{a_i} - p_{b_i}||}$$

(3)

Fig. 3. An illustration of reprojection error from a 2D-3D line correspondence.

References

Zhang et al.: Robust and efficient pose estimation from line correspondences, ACCV 2012.

Mirzaei et al.: Globally optimal pose estimation from line correspondences, ICRA 2011.

Dataset 02

Real Data (Non-Perspective):

Table 1: Average error from the multi-camera systems simulated with 3, 5, 10, 20 and 25 cameras from datasets 01 and 02 respectively.

<table>
<thead>
<tr>
<th>Dataset</th>
<th># of Cameras</th>
</tr>
</thead>
<tbody>
<tr>
<td>01 (1 Error)</td>
<td>0.012</td>
</tr>
<tr>
<td>02 (2 Errors)</td>
<td>0.056</td>
</tr>
</tbody>
</table>

Real Data (Perspective):

Table 2: Comparisons of errors from our algorithm, Zhang et al. and Mirzaei et al. for one camera.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Zhang et al.</th>
<th>Mirzaei et al.</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>0.015</td>
<td>0.024</td>
<td>0.035</td>
</tr>
<tr>
<td>02</td>
<td>0.029</td>
<td>0.047</td>
<td>0.028</td>
</tr>
</tbody>
</table>