Automatic Camera and Range Sensor Calibration using a Single Slot

Authors: Andreas Geiger, Frank Moosmann, Omer Car, and Bernhard Schuster

Presenter: Dor Gabay
An automatic calibration system is proposed such that extrinsic and intrinsic camera parameters are recovered as well as transformation between cameras and range sensors all within one minute.

Proposed calibration requires no user intervention and robust to various lighting conditions.

A single image and range scan suffices for calibration in most scenarios.

Camera-to-range registration method discovers multiple solutions in ambiguous cases.

A Kinect 3D sensor (bottom) and Velodyne HDL–64 laser scanner (top) are used to test calibration method.
Motivation

- Current camera calibration tools, such as the Matlab Camera Calibration Toolbox and OpenCV, are not robust and require manual intervention.
- There are not many thoroughly developed camera-to-range sensor calibration methods.
- Kassir proposed a corner detector which can detect the corners of a single checkerboard within an image while the proposed method detects multiple checkerboards.
- Unnikrishnan has proposed a method using a 3D range sensor where the user marks the calibration pattern in an interactive GUI while the proposed method needs only a single shot and no manual labor.
A single image with multiple checkerboards is used (left image).

The camera calibration method for corner detection is as follows:

- Locate checkerboard corners in the image.
- Refine the candidate pixel corners for sub-pixel accuracy.
- Energy function is used to recover checkerboard structures subject to structural constraints.
- Image correspondences are obtained by sampling affine transformations and maximizing a fitness function.
- Non-linear optimization is used to recover camera parameters.
Instead of using Harris points for localizing junctions, a corner likelihood using two different n by n corner prototypes are utilized: 1) with axis-aligned corners (4), 2) 45 degree corners (4).

- Each prototype has 4 images called A,B,C,D

The corner likelihood, c, at pixel i is determined by finding the maximum over all combinations of prototype flippings:

\[ c = \max(s_1^i, s_2^i, s_1^i, s_2^i) \]

\[ s_1^i = \min(\min(f_A^i, f_B^i) - \mu, \mu - \min(f_C^i, f_D^i)) \]

\[ s_2^i = \min(\mu - \min(f_A^i, f_B^i), \min(f_C^i, f_D^i) - \mu) \]

\[ \mu = 0.25 (f_A^i + f_B^i + f_C^i + f_D^i) \]

\( s_1^i \) denote the likelihood of two possible flippings for prototype i. 
\( f_X^i \) is the filter response for Kernel X and prototype i for particular pixel.
Camera-to-Camera Calibration – Corner Detection (2)

- A list of Corner candidates are found using a conservative non-maxima-suppression (with parameters n_nms and tau_nms).
- We use gradient statistics in a local n by n pixel neighborhood to verify candidates:
  - Compute weighted orientation histogram from Sobel filter response and find two dominant modes using mean shift.
  - Construct a template T for expected gradient strength.
  - The normalized cross correlation of T with second norm of I (gradient strength) and the corner likelihood give corner score (threshold with tau_score to give corner candidates).
Camera-to-Camera Calibration – Sub-Pixel Corner and Orientation Refinement

- Corner locations are refined as follows:
  - The image gradient, \( g_p \), at a neighboring pixel, \( p \), of corner location \( c \) is approximately orthogonal to \( p - c \) giving the following optimization problem with solution to the right:
    \[
    c = \arg \min_{c'} \sum_{p \in N_1(c')} (g_p^T(p - c'))^2 \quad \Rightarrow \quad \text{solution: } c = (\sum_{p \in N_1} g_p g_p^T)^{-1} \sum_{p \in N_1} (g_p g_p^T) p
    \]
  - Where \( N_1 \) is 11 by 11 pixel neighborhood around corner

- The edge orientation is refined as follows:
  - We minimize the error in deviation of the orientation vectors', \( e_1 \) and \( e_2 \), normal with respect to image gradient:
    \[
    e_i = \arg \min_{e'_i} \sum_{p \in M_i} (g_p^T e'_i)^2 \quad \text{s.t. } e'_i e'_i = 1
    \]
  - By setting the derivative of the Lagrangian to zero we solve for the eigenvalues, \( g_p \), by finding them with their respective eigenvectors, \( g_p \):
    \[
    \sum_{p \in M_i} \begin{pmatrix} g_{1p}^T g_{2p}^T \\ g_{2p}^T g_{1p}^T \end{pmatrix} \in \mathbb{R}^{2 \times 2}
    \]
Camera-to-Camera Calibration – Structure Recovery

- We recover Y, the labeling of the respective corner candidates X, by minimizing the energy function:

  \[ E_{\text{corner}} \text{ is the negative number of explained checkerboard corners while } E_{\text{struct}} \text{ measures the ability of two neighboring corners } i \text{ and } j \text{ to predict a third neighbor } k. \]

- An iterative operation around a seed corner (a) is used to search for the closest neighbors in the direction of the edges to minimize \( E(X,Y) \) in order to deduce the final image (c).
Camera-to-Camera Calibration – Matching and Optimization

- After the checkerboards are recovered in all camera images, we determine the corresponding corners using a simple matching algorithm:
  - We use one camera as a reference camera and match all other cameras using a 2D similarity transformation:

\[
\varphi(p; A, b) = \bar{A}p + b.
\]

- Where function accounts for translation (b), scaling (A), and rotation (A).

- We optimize the problem with 10 intrinsic and 6 extrinsic parameters using the Matlab Camera Calibration Toolbox.
We seek to estimate the 6-DOF rigid transformation parameters which specify relative pose of reference camera coordinate system with respect to coordinate system of range sensor:

The proposed algorithm is as follows:
- Segmentation is used to find planes in the range data.
- Transformation Hypothesis are generated by random plane association.
- The chosen planes are refined and verified.
- Non-Maxima-Suppression step is used to yield feasible solutions.
Experimentation

- We use a total of 126 camera images, 55 range measurements, and parameter/threshold/Sobel filter sizes values that are equal for all trials of the experiment in 10 different calibration settings.
- Most time is spent on camera-to-range point cloud optimization procedures.
- The total time of the automatic calibration method is just under 1 minute.
- We determine the performance of the following:
  - Corner Detection and Checkerboard Matching
  - Camera-to-Camera Calibration
  - Camera-to-Range Calibration
Several calibration methods are compared, including Harris Corners, Ha et al., and Kassir methods.

The detection threshold is varied to produce the plots shown on the top and a close up on the bottom.

The proposed method outperforms all baselines, in particular in terms of Recall (this is evident from the plots presented to the left).
We evaluate Camera-to-Camera Calibration by analyzing the focal lengths and ground truth values of all 10 calibration settings shown in a table (next slide).

- Ground truth values may be prone to slight error.

The Poisson Distribution of the reprojection errors is shown to the left. On average, the reprojection error is 0.18 pixels indicating a good model fit.

The errors of the fully automatic calibration system are generally small, but are large for settings 7 and 8 due to outdoor setups making cast shadows difficult to interpret the checkerboard image.
### Experimentation – Camera-to-Camera Calibration (2)

<table>
<thead>
<tr>
<th>Setting</th>
<th># Shots</th>
<th>Range sensor</th>
<th>Environment</th>
<th># Checkerboards</th>
<th>Focal length f</th>
<th>Mean (f)</th>
<th>Std (f)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5</td>
<td>kinect</td>
<td>indoor</td>
<td>3</td>
<td>6.00 mm</td>
<td>6.10 mm</td>
<td>0.01 mm</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>kinect</td>
<td>indoor</td>
<td>3</td>
<td>9.50 mm</td>
<td>9.68 mm</td>
<td>0.13 mm</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>kinect</td>
<td>indoor</td>
<td>3</td>
<td>9.50 mm</td>
<td>9.68 mm</td>
<td>0.01 mm</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>kinect</td>
<td>indoor</td>
<td>3</td>
<td>4.00 mm</td>
<td>4.12 mm</td>
<td>0.01 mm</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>kinect</td>
<td>indoor</td>
<td>3</td>
<td>4.00 mm</td>
<td>4.11 mm</td>
<td>0.01 mm</td>
</tr>
<tr>
<td>5</td>
<td>9</td>
<td>kinect</td>
<td>indoor</td>
<td>3</td>
<td>11.50 mm</td>
<td>11.61 mm</td>
<td>0.19 mm</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>velodyne</td>
<td>hall</td>
<td>7</td>
<td>4.00 mm</td>
<td>4.53 mm</td>
<td>0.11 mm</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>velodyne</td>
<td>hall / outdoor</td>
<td>4–7</td>
<td>8.00 mm</td>
<td>8.25 mm</td>
<td>0.29 mm</td>
</tr>
<tr>
<td>8</td>
<td>6</td>
<td>velodyne</td>
<td>outdoor</td>
<td>4–5</td>
<td>4.00 mm</td>
<td>4.32 mm</td>
<td>0.11 mm</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>velodyne</td>
<td>garage</td>
<td>8–12</td>
<td>4.00 mm</td>
<td>4.23 mm</td>
<td>0.02 mm</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Baseline b</td>
<td>Mean (b)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>11.50 cm</td>
<td>11.45 cm</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>11.50 cm</td>
<td>11.11 cm</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20.70 cm</td>
<td>20.98 cm</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20.70 cm</td>
<td>20.76 cm</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6.20 cm</td>
<td>6.19 cm</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6.20 cm</td>
<td>6.04 cm</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>53.60 cm</td>
<td>52.27 cm</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>53.60 cm</td>
<td>56.24 cm</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>53.60 cm</td>
<td>54.10 cm</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>53.60 cm</td>
<td>53.17 cm</td>
</tr>
</tbody>
</table>

Monocular, wide angle, and fisheye calibrations
We evaluate the robustness of camera-to-range calibration by measuring its result versus ground truth values (prone to error) for a given calibration setting and variations of Gaussian noise.

The measures errors against the ground truth values are measured for both rotation and translation is whisker plots for different noises.

\[
e_t = \|t - t_g\| \\
e_r = \angle(R^{-1}R_g)
\]
The automatic calibration method has been shown to be effective in various lighting conditions.

The limiting assumption of the calibration method is the common field of view of the camera and the range sensors.

Useful for applications of generating stereo of scene flow ground truth and augmenting images with depth values or colorizing a point cloud.

Applications of Calibration method can be broadened if the method could handle overlapping fields of view.
References

Questions

1. Please briefly describe the corner detection method used in the camera-to-camera calibration method.
2. What is the basic process of the camera-to-range calibration?
3. What process in the whole automatic calibration scheme encompasses most of the running time during experimentation?
4. In what calibration setting conditions does the camera-to-camera calibration prove to be less effective?
5. How does the camera-to-range calibration vary to noise?