JOINT CALIBRATION OF A CAMERA TRIplet AND A LASER RANGEFINDER

Florian Schweiger, Ingo Bauermann, Eckehard Steinbach
Technische Universität München
Institute of Communication Networks
Media Technology Group
80290 München

ABSTRACT
This paper studies the calibration of a multi-sensor capture device for the acquisition of image-based scene representations. It is composed of cameras and a laser rangefinder. We use a plane based approach to calibrate the compound system, determining both intrinsic and extrinsic parameters. The novelty of our approach is that the cameras and the scanner are treated jointly. We have tested our method both in simulations and on real data and observe very good calibration results.

Index Terms— Camera Calibration, Sensor Fusion, Image Based Rendering

1. INTRODUCTION
In this paper we present a calibration algorithm for the multi-sensor capture device that is shown in Figure 1. The device can be used to acquire data sets for image based rendering. The cameras capture the visual content of a scene, and they can be used to estimate its 3D structure. The laser rangefinder is intended to improve the behaviour on unstructured surfaces where depth estimation by the cameras alone is problematic. Another advantage is that the rangefinder provides absolute depth values, in contrast to a camera, that cannot determine offhand the true extent of a scene.

Our goal is to determine both the intrinsic and the extrinsic parameters of the components with respect to each other.

1.1. Setup of our SICK-TRIVIS System
The three USB video cameras are structurally identical and they are mounted on a rack so as to form a triangle with a side length of approximately 25 cm. Their lines of sight are not strictly collateral but converge slightly. The resolution of their 1/2° CCD sensor is 1280 × 1024 pixels and their field of view spans approximately 45°. This implies a focal length of around 10 mm. The cameras are synchronized and provide about 10 frames per second each. The laser rangefinder is situated between the cameras and points in the same overall direction. It measures the distance of objects in a distinct plane that is perpendicular to the triangle spanned by the cameras.

Fig. 1. The SICK-TRIVIS capture device composed of three cameras and a laser rangefinder. For illustration, the plane scanned by the rangefinder is overlaid.

and parallel to the baseline joining the lower two cameras. Its field of view is 100° and its resolution 0.25°. According to the manufacturer, the tolerance in depth measurement is ±1 mm. The considered plane is scanned once every 20 ms and the output is synchronized with the cameras.

We choose the coordinate system of the device to coincide with that of the scanner, i.e., the rangefinder resides in the point of origin, and it scans the plane \( z = 0 \), pointing along the \( y \)-axis.

1.2. Related Work
Calibration of camera systems is well understood and there is a vast number of algorithms for this task. One particularly flexible tool is that proposed by Svoboda et al. [1].

Multi-sensor platforms, composed of cameras and a sensor of another type, such as a rangefinder, are typically cal-
ibrated by treating the camera subsystem separately, and by inserting the alien sensor afterwards \[2\] [3]. In this case, of course, information provided by the rangefinder is ignored during camera calibration. We calibrate the complete system jointly in order to benefit from the additional sensor. To our knowledge, no joint calibration method has been proposed for a multi-camera system and a laser rangefinder in the literature so far. Our approach is closely related to plane based calibration in \[4,5,6,7,8\]. The novelty here is that we use a virtual plane that is spanned by the measurements of the rangefinder.

The fact that the cameras’ viewing directions are almost parallel poses a problem to many multi-camera calibration methods. Our algorithm can cope with this critical configuration.

1.3. Overview

This paper is structured as follows. Section 2 introduces notation and presents the theoretical foundations of our approach. In Section 3 we discuss the results obtained by our algorithm before we conclude the paper in Section 4 summarizing and giving an outlook on possible improvements.

2. CALIBRATION APPROACH

2.1. Notation

We use the standard pinhole model for the cameras with projection matrices of the following form:

\[ P_i = KR_i(I_{3\times3} - C_i), \quad i = 1, 2, 3 \] (1)

Here, \( K \) is the \( 3\times3 \) upper triangle matrix containing the intrinsic parameters of the cameras, which we assume identical in first approximation. \( R_i \) is the \( 3\times3 \) rotation matrix representing the orientation of the \( i \)th camera and \( C_i \) is its projection center in inhomogeneous coordinates.

In order to incorporate the laser rangefinder into this framework we model it as an affine camera performing an orthographic projection along the \( z \)-axis. The corresponding projection matrix is given by \( S \), where \( S = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \). It should be mentioned here that the rangefinder can of course not detect points off the plane \( z = 0 \). The above “scanner camera” model is thus only applicable to points on this plane.

A 3D world point in homogeneous representation is denoted by the 4-vector \( \mathbf{X} \). Its images in the \( i \)th camera and in the scanner are then given by \( \mathbf{x}_i = P_i \mathbf{X} \) and \( \mathbf{s} = S \mathbf{X} \).

Our calibration approach uses the Image of the Absolute Conic (IAC) which we will denote by \( \omega = K^{-T} K^{-1} \).

2.2. Plane Based Calibration

The foundation of our approach is the concept of plane induced homographies. It can be shown \[8\] that the image of a set of coplanar world points can be mapped linearly from one camera to another.

Let \( H_i \) be the homography from the scanner image to the \( i \)th camera image. Then, there are two ways of imaging a world point \( \mathbf{X} = (X, Y, 0, 1)^T \), either via this homography or simply by the respective projection matrix.

\[
\mathbf{x}_i = H_i \mathbf{s} = H_i S \mathbf{X} = H_i \begin{pmatrix} X \\ Y \\ 1 \end{pmatrix}
\]

\[
\mathbf{x}_i = P_i \mathbf{X} = \begin{pmatrix} p_{i1} & p_{i2} & p_{i3} \end{pmatrix} \begin{pmatrix} X \\ Y \\ 1 \end{pmatrix}
\]

with \( m^j \) the \( j \)th column of a matrix \( M \).

Since the above equations hold for every point on the scan plane \( z = 0 \) it follows that \( H_i \) is equivalent (equal up to scale) to \( P_i \) with the third column omitted.

\[
P_i = \begin{pmatrix} h_{i1} & h_{i2} & * \\ h_{i2} & h_{i3} & * \\ * & * & * \end{pmatrix}
\] (2)

The \( H_i \) can be determined from measured point correspondences \( \tilde{s}^k \leftrightarrow \hat{x}_k^i \), \( k = 1, \ldots, n \) with \( n \geq 4 \). In particular, we compute initial estimates using the Direct Linear Transform (DLT) algorithm according to \[8\]. Then we refine these to get maximum likelihood estimates under the assumption of independent zero mean Gaussian noise in the measured point coordinates (with assumed variance \( \sigma^2 \) in the camera images and \( \sigma^2 \) in the scanner image). This is achieved by minimizing the following error measure.

\[
\sum_{k=1}^{n} \left( \frac{d(\tilde{s}^k, s^k)^2}{\sigma_s^2} + \sum_{i=1}^{3} \frac{d(\hat{x}_k^i, H_i s^k)^2}{\sigma^2} \right)
\]

Basically, this weighted sum of geometric distances represents the overall reprojection error, both in the scanner and in the camera images. The homographies \( H_i \) and the points \( s^k \) are the variables in this minimization process. The ratio \( \sigma_s : \sigma \) must be specified manually.

Equipped with this consistent set of homographies \( H_i \), we now turn to the retrieval of the missing third columns of the projection matrices \( P_i \). From (1) and (2) it is easily verified that

\[
R_i^T R_i = \begin{pmatrix} h_{i1}^T h_{i1} & h_{i1}^T h_{i2} & h_{i1}^T h_{i3} \\ h_{i2}^T h_{i1} & h_{i2}^T h_{i2} & h_{i2}^T h_{i3} \\ h_{i3}^T h_{i1} & h_{i3}^T h_{i2} & h_{i3}^T h_{i3} \end{pmatrix}
\]

Due to the orthogonality of the \( R_i \) this implies a total of six constraints on the IAC \( \omega = K^{-T} K^{-1} \), namely for \( i = 1, 2, 3 \):

\[
h_{i1}^T \omega h_{i1} = h_{i2}^T \omega h_{i2} = h_{i3}^T \omega h_{i3} = 0
\]

As a homogeneous, symmetric matrix, the IAC has only 5 degrees of freedom. An appropriate parameterization thus leads
to an overdetermined set of equations that can be solved for \( \omega \) in the sense of least squares; \( K \) is subsequently determined using Cholesky factorization.

With knowledge of the cameras’ intrinsic parameters \( K \), the still missing third columns of their respective projection matrix can be computed. Applying \( K^{-1} \) to a camera matrix \( P_i \) yields the partial camera orientation \( R_i \).

\[
R_i = K^{-1} \begin{bmatrix} h_1^i & h_2^i & u^i \\ h_3^i & h_4^i & s^i \end{bmatrix}
\]

The remaining third column can then be computed as the cross product \( \pm (K^{-1}h_1^i) \times (K^{-1}h_2^i) \). The final ambiguity in sign can be resolved considering the relative position of the cameras to the plane \( z = 0 \). The camera centers can, according to \([1]\) and \([2]\), be computed as \( C_i = -R_i^T K^{-1}h_i^i \) and, given the physical setup, the \( z \)-component has to be positive for the upper camera and negative for the lower two. The \( P_i \) can now be assembled according to \([1]\).

Finally, global bundle adjustment is performed that modifies all camera parameters and the reconstructed cloud of world points. It is during this step that points off the plane \( z = 0 \), \textit{i.e.}, those undetected by the rangefinder, are incorporated into the calibration process. We use the Levenberg-Marquart algorithm.

### 3. RESULTS

#### 3.1. Experiments on Synthetic Data

First experiments were conducted in a virtual environment. Therefore, we analytically defined a capture device simulating the SICK-TRIVIS setup. A set of 140 randomly generated world points, 40 of which lie on the plane \( z = 0 \), were then imaged and zero mean Gaussian noise of varying variance was added to the so obtained image point coordinates. Noise variance \( \sigma^2_s \) is in \( \text{mm}^2 \) for the scanner and was chosen 9 times higher than the variance \( \sigma^2 \) in pixels for the camera images.

The experiments revealed that all the camera parameters can be accurately recovered. Figure 2 exemplarily shows the relative error in focal length. For reasonable noise levels it stays below 0.5\%. Furthermore, the 3D setup can be recovered with a mean absolute error of about 10 mm. Given the working volume spanning several cubic meters, this is acceptable. The obtained values of reprojection error are plotted in Figure 3.

#### 3.2. Experiments on Real Data

In order to acquire point correspondences between the cameras and the rangefinder we extended an approach presented in \([1]\) where a laser pointer is waved in front of the cameras yielding easily detectable point correspondences. In order to make the tip of the laser pointer visible to the scanner, we stuck it through a piece of cardboard, approximately \( 30 \times 20 \text{cm}^2 \) in size. When this primitive compound object is carefully moved through the scan plane the laser pointer leaves a characteristic peak (see Fig. 4). The main purpose of the cardboard is to constitute a background for this peak. The laser pointer tip itself, without cardboard, is difficult to identify. Moreover, if the depth difference between pointer tip and background is too important, bothersome ringing artefacts occur, which confine the exact detectability of the laser pointer even further.

We then extracted the peak locations as well as the pixel coordinates of the laser point manually. Altogether, we captured 20 points that were seen by all three cameras and another 19 points that were also visible to the rangefinder.

The algorithm of Section 2.2 was applied yielding the results in Table 1. It shows the obtained reprojection errors after only the 19 coplanar points were processed, and the final results after bundle adjustment with all measurements contributing. The noise ratio \( \sigma_s : \sigma \) was chosen somewhat heuristically to be \( 3 \text{ mm} : 1 \text{ px} \). This value agrees with the assumed accuracy in the manual extraction of points from the scanner and camera images. In accordance with our experiments on synthetic data, we achieve low reprojection error that is well balanced between the sensors. The reprojection error in the cameras is similar to what can be achieved by multi-camera calibration algorithms, such as \([1]\), with the plus of the scanner already being integrated.

#### 4. CONCLUSION

In this paper, we have presented a joint calibration method for our SICK-TRIVIS system. Although the given setup is close to degenerate, our algorithm yields good results. In terms of reprojection error it can compete with other calibration methods that treat the cameras separately.
Fig. 3. Reprojection error over noise standard deviation $\sigma$ in pixels. The graph shows the mean squared error averaged over 100 runs per evaluated noise level. Only camera 1 is plotted exemplarily. The unit is $\text{px}^2$ for the camera and $\text{cm}^2$ for the scanner.

<table>
<thead>
<tr>
<th>camera 1</th>
<th>before bundle adj.</th>
<th>after bundle adj.</th>
</tr>
</thead>
<tbody>
<tr>
<td>camera 1</td>
<td>1.6514 $\text{px}^2$</td>
<td>0.52517 $\text{px}^2$</td>
</tr>
<tr>
<td>camera 2</td>
<td>3.2853 $\text{px}^2$</td>
<td>0.41527 $\text{px}^2$</td>
</tr>
<tr>
<td>camera 3</td>
<td>3.8314 $\text{px}^2$</td>
<td>0.53668 $\text{px}^2$</td>
</tr>
<tr>
<td>scanner</td>
<td>166.4275 $\text{mm}^2$</td>
<td>16.9316 $\text{mm}^2$</td>
</tr>
</tbody>
</table>

Table 1. Reprojection errors before and after bundle adjustment.

A first open problem to be addressed is the estimation or the prediction of the relative noise variances in scanner and cameras, which is required for optimal results. It should also be investigated how the process of point extraction could be automatized in order to have a larger number of point correspondences available, and so to gain robustness.

5. REFERENCES


