Modeling and Calibration of Coupled Fish-Eye CCD Camera and Laser Range Scanner for Outdoor Environment Reconstruction

Xavier Brun, Francois Goulette

To cite this version:

Xavier Brun, Francois Goulette. Modeling and Calibration of Coupled Fish-Eye CCD Camera and Laser Range Scanner for Outdoor Environment Reconstruction. 3-D Digital Imaging and Modeling, 2007. 3DIM’07, 2007, Montréal, Québec, Canada. hal-01259676

HAL Id: hal-01259676
https://hal.science/hal-01259676
Submitted on 21 Jan 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Modeling and Calibration of Coupled Fish-Eye CCD Camera and Laser Range Scanner for Outdoor Environment Reconstruction

Xavier Brun  
Mines Paris  
75272 Paris Cedex 06, France  
brun@ensmp.fr

François Goulette  
Mines Paris  
75272 Paris Cedex 06, France  
goulette@ensmp.fr

Abstract

Precise and realistic models of outdoor environments such as cities and roads are useful for various applications. In order to do so, geometry and photography of environments must be captured. We present in this paper a coupled system, based on a fish-eye lens CCD camera and a laser range scanner, aimed at capturing color and geometry in this context. To use this system, a relevant model and an accurate calibration method are presented. The calibration method uses a simplified fish-eye model; the method uses only one image for fish-eye parameters, and avoids the use of large calibration patterns as required in others methods. The validity and precision of the method are assessed and example of colored 3D points produced by the system is presented.

1. Introduction

Models of outdoor environments such as cities and roads are useful for various applications, such as architecture planning, 3D cartography for car and pedestrian navigation, virtual tourism, Virtual Reality, video games, and so on. A noticeable example is the recent use of the third dimension in popular cartography services on the Internet, such as "Virtual Earth" (Microsoft©) or "Google Earth©", on which one may see 3D model details of the size of buildings.

In order to produce realistic modeling of outdoor environments, a precise digitizing of 3D geometries and textures is needed, at the level and scale of the desired use (e.g. ground level for car and pedestrian navigation). To perform this work, a new generation of adapted scanning devices, named Mobile Mapping Systems, are appearing. For 3D geometries, one can use passive methods such as photogrammetry or stereovision to deliver the location of specific points or features in the scene, or active methods using lasers or structured light to deliver 3D points all over a scanned surface. Active methods usually result in more dense data sampling and are quite adapted to capture directly geometry. For these reasons, several people have explored the use of laser range sensors mounted on vehicles. One possibility is to define several scanning spots (stop and go), and to make a fusion of the 3D data clouds with adapted frame transformation between the locations [1, 2, 3, 17, 16]. Another possibility is to use the vehicle itself as a scanning device, its movement defining one of the scanning directions [4, 5, 6, 7].

Figure 1. Urban scene, range scanning alone

To add textures, additional cameras may be added to the mobile mapping system, and the information are combined during post-processing [1, 4]. Geometry and texture data being obtained with different sensors with various characteristics, it is of importance to set up a correct model of these sensors, and to perform through calibration a correct identification of its numerical parameters.

We present in this paper the use of a system coupling the capture of geometry and texture for outdoor environment reconstruction with a Mobile Mapping System. The sys-
tem is based on a fish-eye CCD camera and a laser range scanner. The fish-eye camera allows to capture very wide angles and to limit the number of cameras needed when regular lenses are used. However in that case classical camera models are not adequate when fish-eye lenses are needed, and one has to use specific models and calibration methods. We present the model used for our application in the section 3, that is of intermediate complexity, and a simplified calibration method. We discuss the validity and precision of our method, using for reasoning a sensitivity analysis of the parameters, and an experimental test in the section 4. Results are presented for the mobile mapping system we are using.

2. Coupling geometry and color

We are using a Mobile Mapping System consisting of a car equipped with localization sensors (GPS, IMU) and a laser range scanner. The scanner is placed at the rear of the vehicle, and it performs scanning over a vertical plane perpendicular to the direction of the car. During the movement of the car, at low or normal speed, the laser is thus scanning the whole environment driven through. Typical numerical values in the context of acquisitions are given in Table 1 and a result of scanning in urban environment is presented on Figure 1.

<table>
<thead>
<tr>
<th>Vehicle</th>
<th>Speed</th>
<th>20 km/h</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buildings</td>
<td>Distances</td>
<td>5-10 m</td>
</tr>
<tr>
<td>Ground, road</td>
<td>Distances</td>
<td>1.5-5 m</td>
</tr>
</tbody>
</table>

The system provides 3D points and facets. We are interested in adding color to the 3D points and texture to the facets. In order to do so, we considered various possibilities based on the existing system, adding cameras to the platform supporting the scanner. We focused on the use of commercial off-the-shelf elements, namely CCD cameras and commercial lenses. In order to produce photometric redundancy for texture map creation further in the processing, we have chosen matrixial CCD that give a complete scene area for each frame (differently from linear CCD). As data capture must be done during the movement of the vehicle, we need to have a rate of several frames per second. The resolution chosen is typical of usual products (Table 2).

The range scanner covers an area of 270°. We present in Table 2 a comparison of angles that would be covered by our camera with three lenses with different focal lengths. Fish-eye lenses are specific lenses using different optics principles than regular ones, that allow to obtain very short focal length and then to cover up to more than 180° in all directions. Using a camera with fish-eye lens allows to capture a complete half-space with only one camera. In that case, the resulting image fit in a disk which is inscribed inside the rectangular CCD array; This lead to have the same aperture angle in horizontal and vertical directions. To cover the same area as the scanner, only two cameras with fish-eye lenses would be needed, whereas in the other cases presented in Table 2 we would require 3 cameras with the wide-angle lens and 5 with the regular one. This motivated us to explore the use of a coupled system using laser range sensor for geometry, and fish-eye lens mounted on CCD camera for images (Figure 2).

<table>
<thead>
<tr>
<th>Laser scanner</th>
<th>Number of points</th>
<th>1080 per profile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Profiles rate</td>
<td>270°</td>
<td></td>
</tr>
<tr>
<td>Color CCD</td>
<td>Number of pixels</td>
<td>780x582 pixels</td>
</tr>
<tr>
<td>CCD size (1/2&quot;)</td>
<td>6.47mm x 4.83mm</td>
<td></td>
</tr>
<tr>
<td>Frames rate</td>
<td>10 Hz</td>
<td></td>
</tr>
<tr>
<td>Regular lens</td>
<td>Focal length</td>
<td>12 mm</td>
</tr>
<tr>
<td>Aperture angle</td>
<td>57° × 44°</td>
<td></td>
</tr>
<tr>
<td>Wide-angle</td>
<td>Focal length</td>
<td>4.2 mm</td>
</tr>
<tr>
<td>Aperture angle</td>
<td>114° × 98°</td>
<td></td>
</tr>
<tr>
<td>Fish-eye lens</td>
<td>Focal length</td>
<td>1.4 mm</td>
</tr>
<tr>
<td>Aperture angle</td>
<td>185° × 185°</td>
<td></td>
</tr>
</tbody>
</table>

3. Modeling and calibration

The system presented above is, as it where, an integrated sensor, giving measures of geometry and color of objects in its surrounding environment. In order to be able to use these measures, one has to know a model linking the reality to its perception, which is the combination of a mathematical model and the numerical values of its parameters. We present in the following the model we use for our sensor and the calibration procedure to obtain its parameters.
Usually for mobile perception sensors, one makes the distinction between intrinsic and extrinsic models (or parameters). The intrinsic model gives the relationship between the observed reality and its measures, in a frame relative to the sensor, whereas the extrinsic model gives the transformation between the relative frame of the sensor to a fixed frame. This distinction applies to both the camera and the laser range scanner. However, as they are fixed together on a common support, if we assume that the support is rigid enough, we can say that the position and orientation of one of the elements, and the rigid transformation between this element and the second one, is enough to determine the two extrinsic parameters. As the common support is fixed on a geo-referenced vehicle, there remains only three models/parameters to determine:

1. Rigid camera-scanner frame transformation
2. Intrinsic laser scanner model
3. Intrinsic camera model

We present below these models and the calibration methods we used. Concerning the intrinsic camera model, the classical pin-hole model is no longer appropriate for fisheye lenses. We found that the models and methods of the literature were not well suited in our case, so we present a simplified new model and its corresponding calibration method that we developed for this application.

### 3.1. Laser scanner model

The intrinsic laser scanner model is indeed straightforward. The scanner returns profiles which are series of distances scanned with a constant angle step around an axis. The position of a measure in the profile gives it angle. Hence the coordinates of the corresponding point, in the scanning plane of the scanner, are:

\[
P_l = d_l \times \left( 0 \quad \sin(\theta_l) \quad \cos(\theta_l) \right)
\]

As far as we assume that the geometry of acquisition is correct and that the inside firmware is well tuned, this model does not require calibration.

### 3.2. Camera-scanner frame change

The scanner frame consists of the scanning plane and rotation axis. It is centered on the center of rotation; there is an offset angle in the scanning plane. The camera frame consists of the CCD plane and the optical axis, that are supposed to be orthogonal. The center of the frame is at the intersection between CCD plane and optical axis.

The calibration aims at determining the rigid transformation between camera and scanner frames, represented as a rotation matrix \( \Phi \) and a translation vector \( \Delta \). If we denote \( P_c \) the vector coordinates of a point in camera frame and \( P_l \) the coordinates of the same point in scanner frame, the rigid transformation states that for each point:

\[
P_l = \Phi P_c + \Delta
\]

To perform the calibration, we use the method described by Pless and Zhang [8] and improved by [9]. This method uses a planar calibration pattern (chessboard) that is viewed by the camera and the scanner at the same time (Figure 3), in various configurations (positions).

For each configuration, the camera image of the pattern provides 2D image points coming from 3D points that we know lie on a common plane in space, while the laser measurements give 3D points that lie also on this plane. Hence, each configuration of the pattern gives a set of geometric constraints, the rigid transformation must put those laser control points on the plane define by the camera measurements. A set of several configurations, Pless and Zhang [8] have shown at least fifteen are needed, allows to determine the unknown parameters.

We define an error function based on the reprojection errors of the laser points for all configurations. Denoting, for the \( i \)th configuration, \( N_i \) is the normal vector of the plane and \( d_i \) its distance to the origin, and \( P_{ij} \) the laser points, it is expressed as:

\[
\sum_i \sum_j \left( \frac{N_i}{\|N_i\|} (\Phi^{-1} P_{ij} - \Delta) - d_i \right)^2
\]

The solution parameters are obtained with the Levenberg Marquardt non-linear optimization method. An algebraic solution is used as initial guess [8] and the algorithm converges in a few (10-20) iterations.

We perform the calibration with a regular lens placed on the camera, because we need, for each configuration, to use at first the Tsai method [10] to get the camera position in the world reference system defined by the pattern. As far as the precision of the extrinsic parameter calibration depends on the precision of location of points of a pattern in the camera, we assume that doing so can only provide better
results, because for a given resolution, the precision of observation in the camera is better with a larger focal length. After calibration with regular lens, changing to the fish-eye lens possibly affects only the position of the optical axis, which is taken into account as parameters in the intrinsic fish-eye model.

3.3. Fish-eye camera model

3.3.1. Model used

The usual camera models found in the literature (perspective or pin-hole, pin-hole with distortions) are not appropriate any longer for fisheye lenses. Indeed, with that kind of wide-angle lens, the Gauss conditions of optics (light quasi-parallel to the optics axis and touching the lens close to its center) are no longer available - lenses have aperture angles up to 180° (Figure 4). More appropriate models for fish-eyes lenses are equidistance or generic projection models. If we denote θ the angle between the optical axis and the line relating the optical axis and the object viewed (Figure 5), and r the distance of the projection of the point on the CCD plane, these models give the relationship between r and θ in various forms:

\[ r(\theta) = f \tan(\theta) \quad \text{perspective} \]
\[ r(\theta) = f\theta \quad \text{equidistance} \]
\[ r(\theta) = k_1\theta + k_2\theta^2 + k_3\theta^3 + \ldots \quad \text{generic} \]

After this projection, one has to transform further into pixel units. Four parameters are needed, \((u_0, v_0)\) the principal point, and \(\mu_x, \mu_y\) the linear densities of pixels in horizontal and vertical directions. For a 3D point \(P(\rho, \theta, \varphi)\), the pixel values in the image are:

\[
\begin{pmatrix}
  u \\
  v
\end{pmatrix} = \begin{pmatrix}
  u_0 \\
  v_0
\end{pmatrix} + \begin{pmatrix}
  \mu_x & 0 \\
  0 & \mu_y
\end{pmatrix} r(\theta) \begin{pmatrix}
  \cos(\varphi) \\
  \sin(\varphi)
\end{pmatrix}
\]

Ishii [11] uses the equidistance model without estimating the four parameters \((u_0, v_0, \mu_x, \mu_y)\) to get the pixel values: the principal point is assumed to be on the center of the image \((u_0, v_0) = (\text{width}/2, \text{height}/2)\) and the pixel cells are supposed to be squared \((\mu_x = \mu_y)\). In this model, only one parameter is needed, the focal length expressed in pixel units \(f\mu\). It can be estimated with the maximal value of \(r\) and \(\theta\). \(r_{\text{max}}\) is equal to half the height of the picture and \(\theta_{\text{max}}\) can be estimated by the constructor value of the aperture angle. Ishii use this model to determine \(\theta\) and to reproject the point according to the perspective model, to undistort fisheye image. Alternatively, Kannala [12] has proposed a way to calibrate the generic model and the others parameters as a classical camera calibration process. The disadvantage of his method is the necessity of a very big calibration pattern \((5 \times 2 \, m)\). In practice, such a big and accurate chessboard pattern is expensive.

For our needs, we propose to use an intermediate model between those two (Ishii and Kannala), with three parameters: it consists of the equidistance projection model with the focal length in pixel units \(f\mu\) (pixels are considered squared), and additionally the two parameters of the principal point \((u_0, v_0)\). The model is just a little more developed than Ishii’s one, but simplified from Kannala’s, and it is much easier to calibrate and enough for our needs.

3.3.2. Calibration method

The three parameters \((f\mu, u_0, v_0)\) of our model can be determined easily, without calibration pattern, simply considering an ordinary fish-eye image such as the one on Figure 4.

As pixels are squared, we assume the world projection on the CCD sensor to be a circle. Others pixels are black which allow us to determinate easily which pixel belongs to the circle. We simply use different classical image processing methods to estimate the parameters of the circle, that will be enough for the calibration. First of all, extraction of the circle border is performed with a Canny filter. Afterwards, the position of the center and the radius of the circle may be determined with various methods:
• Barycenter of the no-black (internal) pixels,
• Best fitting circle by mean squared estimation
• Hough circular transformation [13]

The center of the circle gives the position of the optical center projection on the image \((u_0, v_0)\). The circle radius \(r_c\), already in pixel units, is used to estimate the \(f\mu\) parameter the following way: for all pixels, the distance \(r\) between pixel position and center of the image is proportional to the angle \(\theta\) \((r = f\theta, \text{see Figure 5})\). This is also true for extremal values, and \(\theta_{\text{max}}\) is half of the aperture angle of the fisheye (given by the manufacturer). This leads to the result:

\[
(f\mu) = \frac{r_c}{\theta_{\text{max}}}.
\]

This method is simple and robust. We have observed that using different, random-kind of images, does not affect the numerical results. Moreover, if the circle is truncated (which can happen for different camera-lens parameters), it still works (determination of radius of circle). It can also be extended to non-squared pixels, replacing the determination of circle by ellipse. Let us notice that the cell density parameters \((\mu)\) is given by the Tsai camera calibration, but we do not need them as far as the value of focal length is directly determined in pixels units.

4. Precision and validity of the results

We intend to use the coupled laser-point system for two applications. The first one is to enhance the cloud of 3D points with colors, displaying colored points instead of simple black or white dots such as in Figure 1. For this application, the algorithm imagined is to find, for which 3D point obtained with the laser scanner, its corresponding color pixel in the camera image. The maximum precision that can be reached is to hit the correct pixel, which means that the color assigned to the point is the valid one. 

In the following, we present two studies performed on our coupled system. The first one is a sensitivity analysis, which confirms the pertinence of the simplified fish-eye model. The second one is an experimental study, giving numerical results to the two criteria defined above and confirming numerically the quality of the overall procedure.

4.1. Sensitivity analysis

We have performed sensitivity analysis on the overall model of our coupled system (scanner and camera), using the generic model for fisheye projection (Equation (4)), to figure out the relative influence of the parameters used during the projection of a point from laser coordinates to image coordinates. This gives the possibility to consider whether neglecting the third order for fish-eye lens, as described above, is relevant.

Sensitivity analysis gives information on the relative impact of precision on the parameters of a function, to its results. A sensitivity close to zero does not mean that its corresponding parameter is negligible, but that if we change a little its value, the result will be affected only a little. Any interested reader non familiar with sensitivity analysis can find an introduction in [14]). Considering a function \(f(X, p)\), with \(X\) being the vector of variables, and \(p\) the vector of parameters, for each parameter \(p_i\), the sensitivity \(s\) of the function to this parameter is the normed ratio:

\[
s = \frac{\partial f(X, p)}{\partial p_i} \frac{p_i}{f(X, p)}
\]  

In the transformation (projection) from laser coordinates to image, there are six independent parameters for the rigid frame change (three for rotation and three for translation), and six for fish-eye projection \((k_1, k_2, u_0, v_0, x_0, y_0)\). The result of our transformation being a position (projection) in the image, we choose as criterion function \(f(X, p)\) for sensitivity analysis the distance of this position to the center of the image:

\[
f(X, p) = \sqrt{u^2 + v^2}
\]

4.1.1. Estimating numerical sensitivities

Derivatives of the criterion function are determined analytically; sensitivities are then determined numerically with values of actual use. For this estimation, we got several laser points experimentally by scanning a building and reprojecting them into the image, in a context close to our application (3200 points, from several profiles such as depicted on Figure 6). Sensitivies are then computed for each parameter.

Sensitivities are represented in two groups, the group of rigid transformation parameters (9 rotations - only 6 independent, and 3 translations) (Figure 7) and the group of camera parameters (6 parameters) (Figure 8). Sensitivities

\[
\]
are on the Y axis, drawn along the angle $\theta$ of observation from the camera (see Figure 6). The non-visible curve are superimpose in the bottom of the graph.

We can draw several observations of interest from these diagrams. The interest is to compare the numerical values obtained for camera parameters, to those for frame change parameters. One may notice that $(u_0, v_0)$ parameters (center of the camera) have very high sensitivity. This confirms the need to introduce these parameters in the model, because they are not negligible at all. Secondly, one may notice that the $k_2$ parameter corresponding to the third order of the fish-eye model, has a very low sensitivity, compared to all others camera parameters and to frame change parameters. This induces that the third order of the fish-eye model has a low impact on precision, validating the relevance of the model proposed in our approach. To study further this relevance, we present below a numerical estimation of the systematic error induced by neglecting the third-order term of the model.

4.1.2. Influence of third-order term in lens model

The systematic error due to neglecting the third order ($k_2 \theta^3$) term in the fish-eye lens model, have been computed and are presented, in pixel units, on Figure 9 ($\rho$ and $\varphi$ - polar coordinates - being constant, other numerical values coming from earlier calibration results). One can notice that for $\theta$ less than 53° the difference is sub-pixelic. For our application, buildings are typically 10 m away from the camera; given that the camera is at 2 m above the ground, a point needs to be at 16 m high in order to reach an error of 1 pixel due to neglecting the third order of the model. Table 3 shows the correspondance between height and systematic error. As can be seen, considering typical heights of buildings (except in skyscrapers area, out or range anyway for our system), the choice to neglect the third order generates a quite small systematic error.
Table 3. Difference between the equidistance and the generic projection model in pixels according to the object size

<table>
<thead>
<tr>
<th>Angle interval</th>
<th>Height</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>0° – 53.75°</td>
<td>16 m</td>
<td>&lt; 1 pixel</td>
</tr>
<tr>
<td>53.75° – 67.75°</td>
<td>27 m</td>
<td>&lt; 2 pixels</td>
</tr>
<tr>
<td>67.75° – 77.5°</td>
<td>47 m</td>
<td>&lt; 3 pixels</td>
</tr>
<tr>
<td>77.5° – 85.375°</td>
<td>123 m</td>
<td>&lt; 4 pixels</td>
</tr>
</tbody>
</table>

4.2. Experimental Analysis

4.2.1. Experimental Protocol

To evaluate the overall precision of the calibration method presented, we need to use the scanning system on an object for which the geometric and photogrammetric features are well known. We have realized a validation pattern (Figure 10), which allows us to identify easily two sets in the laser data (difference of depth), and for which the color can be identified. This way, it is possible to identify the adequation between geometry and color and therefore to evaluate the precision of calibration. In two series of acquisitions, we got several measurement points on the sets, up to 800 and 1000

![Figure 10. Validation Pattern](image)

4.2.2. Results

The result of the global transformation is a position in the fisheye image so we express the precision in pixel unit. We have presented in the previous section an order of magnitude in meter in the world reference for a pixel according to the camera used. This order of magnitude is still valid in this case. All the points which are not colored correctly belong to borders of each set (in the laser data and in the fisheye image). For each laser point, it is possible to evaluate, if it is not correctly colored, the difference in pixels to the correct color it should have. For the series of points, it is possible to evaluate statistics of errors to the correct projection position in the image (Table 4). The position difference between the sets equal to the precision of the laser scanner (5 cm), which is the worst configuration for the scanning system.

<table>
<thead>
<tr>
<th>Error</th>
<th>Set 1: 1078 points</th>
<th>Set 1: 853 points</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 pixels</td>
<td>96.81%</td>
<td>96.9%</td>
</tr>
<tr>
<td>1 pixels</td>
<td>1.9%</td>
<td>2.1%</td>
</tr>
<tr>
<td>2 pixels</td>
<td>1.2%</td>
<td>0.9%</td>
</tr>
<tr>
<td>3 pixels</td>
<td>0.09%</td>
<td>0.1%</td>
</tr>
</tbody>
</table>

If we consider, as a first order approximation, that the probability distribution of stochastic noise for the results is a Normal (Gaussian), from the results of Table 4 it is possible to draw an estimate of the standard deviation $\sigma$ of the noise (thanks to Normal distributions tables). Here the estimation says that the global precision of the transformation is sub-pixelic:

$$\sigma = 0.53 \text{ pixel}$$

4.2.3. Embedded System

Results of colored 3D points obtained with our system, detailed elsewhere [15], are shown Figure 11 (same scene, with color added, as Figure 1).

![Figure 11. Result of the fusion of the laser range data and fisheye image](image)
are white. As we explain earlier, the resolution of the camera we used allows to get a pixel which cover a 20cm$^2$ area in the world. Hence, first laser points above the vehicle are reprojected with a difference of one or two pixels, explaining the result.

5. Conclusion and discussion

In order to add color and texture to geometries obtained by a laser range scanner geo-referenced on a mobile platform, we have set up a coupled system including a fish-eye CCD camera. We have motivated our choices, and presented in this paper a method for precise, easy and adapted calibration of the coupled system. A precision analysis has shown the relevance of our choices for our application, and has given numerical results that are of good quality for our application. Of interest in our approach, is the possibility to use a simplified model for fish-eye modeling, and a calibration method that does not require, as a difference to other methods, a large calibration pattern.

Ongoing work includes adding a second camera, in order to cover the whole aperture angle of the laser range scanner, and improving the resolution of cameras.

Acknowledgements

We gratefully acknowledge the helpful proofreading comments of Pr. Francis Schmitt for this paper.

References