Precise Omnidirectional Camera Calibration

Dennis Strelow, Jeffrey Mishler, David Koes, and Sanjiv Singh
Carnegie Mellon University
{dstrelow, jmishler, dkoes, ssingh}@cs.cmu.edu

Abstract

Recent omnidirectional camera designs aim a conventional camera at a mirror that expands the camera's field of view. This wide view is ideal for three-dimensional vision tasks such as motion estimation and obstacle detection, but these applications require an accurate model of the imaging process. We present a full model of the imaging process, which includes the rotation and translation between the camera and mirror, and an algorithm that determines this relative position from observations of known points in a single image. We present tests of the model and of the calibration procedure for various amounts of misalignment between the mirror and camera. These tests show that the algorithm recovers the correct relative position, and that by using the full model, accurate shape-from-motion and stereo matching are possible even if the camera and mirror are severely misaligned.

1. Introduction

Recent omnidirectional camera designs aim a conventional camera at a mirror that expands the camera's field of view. This wide view is ideal for three-dimensional vision tasks such as motion estimation and obstacle detection, but these applications require an accurate model of the imaging process. We present a full model of the imaging process, which includes the rotation and translation between the camera and mirror, and an algorithm that determines this relative position from observations of known points in a single image.

We include tests of the model and of the calibration procedure for various degrees of misalignment between the mirror and camera. These tests show that the method recovers the correct relative position, that the method is not unduly sensitive to small random errors in the calibration image observations, and that one image is sufficient to obtain an accurate calibration. The tests also show that by using the full model, accurate shape-from-motion and stereo

matching are possible even if the camera and mirror are severely misaligned.

Previous methods for omnidirectional camera calibration have required that the mirror and camera combination be single viewpoint, i.e., that rays reflected by the mirror onto the camera's center also intersect at a single point inside the mirror. Our calibration is performed within a general minimization framework, and easily accommodates any combination of mirror and camera. In particular, the mirror and camera may be a non-single-viewpoint combination, such as the camera described by Chahl and Srinivasan[1].

For single viewpoint combinations, the advantages of the single viewpoint can be exploited only if the camera and mirror are assumed to be properly aligned. So for these combinations, the simpler single viewpoint projection model, rather than the full model described here, should be adopted if the misalignment between the mirror and camera is sufficiently small. In this case, the calibration algorithm that we describe is useful as a software verification of the alignment accuracy.

2. Related work

Recently, researchers have been highly active in the design of omnidirectional cameras and their application. Nayar's design[6], which combines an orthographic camera with a parabolic mirror, is the most well known. This camera achieves a single viewpoint, i.e., rays that are reflected from the mirror onto the sensor all intersect in a single point inside the mirror. This property allows portions of the omnidirectional image to be remapped into conventional perspective images, allows the image of a world point to be computed easily, and produces circular epipolar curves. However, this system has lower resolution in the center of the image than in the periphery. An alternative design that has approximately uniform resolution is described by Chahl and Srinivasan[1], but as mentioned in the introduction, this combination sacrifices the single viewpoint property. The camera used in our experiments is a refinement of Chahl and Srinivasan's design that provides exact uniformity[7].

There has already been some work on calibration for single viewpoint cameras. Geyer and Daniilidis[2] present a geometric method using two or more sets of parallel lines in one image to determine the camera aspect ratio; a scale factor that is the product of the camera and mirror focal lengths; and an image center, which can be taken to be the (x, y) location of the mirror focal point. Kang[4] describes two methods. The first recovers the image center and mirror parabolic parameter from the image of the mirror's circular boundary in one image; of course, this method requires that the mirror's boundary be visible in the image. The second method uses minimization to recover skew in addition to Geyer and Daniilidis's parameters. In this method the image measurements are point correspondences in multiple image pairs. Because the advantages of single viewpoint cameras are only achieved if the mirror axis is parallel to the camera axis, these methods assume that these axes are parallel rather than determining the relative rotation between the mirror and camera.

The experiments presented in subsections 4.3 and 4.4 below describe the effect of our calibration model on the accuracy of shape-from-motion and estimated epipolar curves, respectively. Some fundamental work on using single viewpoint cameras for these tasks includes Gluckman and Nayer[3], which extended three algorithms for ego-motion estimation with conventional cameras to the single viewpoint omnidirectional case; and Svoboda and Hlavac[10], which describes the epipolar geometry for single viewpoint cameras. Our own previous work on shape-from-motion for non-single-viewpoint cameras is described in [9].

3. Method

3.1 Overview

Our projection model and calibration algorithm separate the conventional camera intrinsics (e.g., focal length, radial distortion) from the relative position between the mirror and camera (i.e., the camera-to-mirror coordinate transformation). The conventional camera intrinsics can be determined using any existing method; for the experiments described here, we have used the method implemented in [5].

Once the camera intrinsics are known, the camera-to-mirror transformation can be determined by obtaining an image of calibration targets whose three-dimensional positions are known, and then minimizing the difference between the observed locations of those targets in an image and the predicted reprojections, with respect to the camera-to-mirror transformation. Figure 1 shows two examples of calibration images used in our experiments. The locations of the three-dimensional points have been surveyed and are known with an accuracy of about two centimeters.

In the following subsections we describe the reprojection of points in the camera coordinate system assuming a general rotation and translation between the camera and mirror (3.2) and the error function we minimize to determine the camera-to-mirror transformation (3.4). In section 3.4 we also briefly describe our equiangular camera design and the effect of misalignment on the projections this camera produces.

3.2 Projection model

Given a camera-to-mirror transformation, computing the projection of a three-dimensional point with a general mirror and camera combination reduces to finding the height z and azimuth θ of the point m on the mirror at which the angle of incidence (i.e., the angle to the three-dimensional point p) equals the angle of reflection (i.e., the angle to the camera's center c). These constraints produce a system of two nonlinear equations in z and θ which we solve numerically:

$$\frac{b_1 \cdot dp}{b_3 \cdot dp} = \frac{-b_1 \cdot dc}{b_3 \cdot dc} \qquad \frac{b_2 \cdot dp}{b_3 \cdot dp} = \frac{-b_2 \cdot dc}{b_3 \cdot dc} \qquad (1)$$

Here, $B = \{b_1, b_2, b_3\}$ is an orthogonal basis for a local coordinate system on the mirror, relative to the global mirror coordinate system. B is centered on the mirror point $m = m(z, \theta)$, with b_1 and b_2 spanning a plane tangent to the mirror at m. dp and dc are the directions from m to the three-dimensional point p and the camera center c, respectively, expressed in the global mirror coordinate system. B, dp, and dc are all functions of the mirror point m. Once m has been determined, it can be transformed from the mirror coordinate system to the camera coordinate system and its projection determined using perspective projection.

The *derivative* of a projection with respect to the components (x_p, y_p, z_p) of the three-dimensional point p is also required by the calibration algorithm. Differentiating the equations (1) in z and θ with respect to x_p, y_p, and z_p produces three 2×2 linear systems that determine columns of the derivative:

$$\frac{\partial m}{\partial p} = \left[\begin{array}{ccc} \frac{\partial z}{\partial x_p} & \frac{\partial z}{\partial y_p} & \frac{\partial z}{\partial z_p} \\ \frac{\partial \theta}{\partial x_p} & \frac{\partial \theta}{\partial y_p} & \frac{\partial \theta}{\partial z_p} \end{array} \right] \qquad (2)$$

With these derivatives in hand the derivatives of the projection with respect to p can be found using the chain rule. The calibration algorithm also requires the derivatives of a projection with respect to the camera-to-mirror parameters, which we approximate using finite differences.

If the mirror and camera axes are assumed to be the same, then (1) reduces to the one-dimensional problem of determining the height z of the mirror point m, which can be solved more efficiently and robustly than the more general problem (1). For single viewpoint mirrors, this simpler
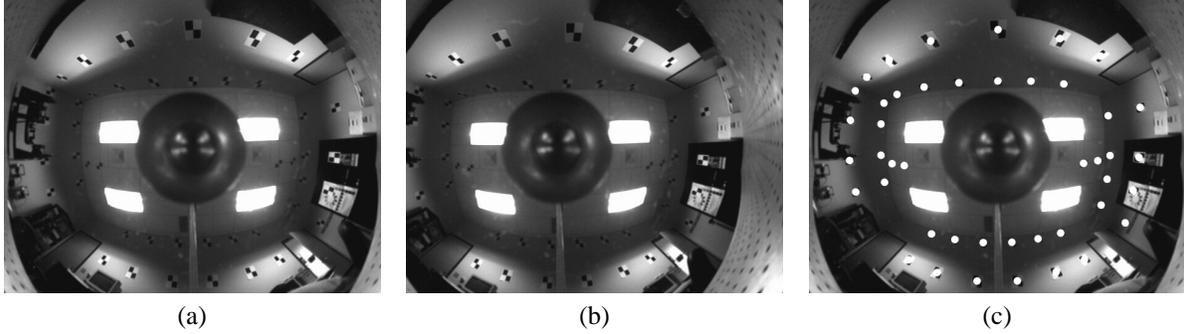
Figure 1. The calibration images for two of our test configrations are shown in (a) and (b). In (a), the mirror and camera are closely aligned. In (b), the mirror and camera are severely misaligned; this is most easily seen by comparing the left and right edges of the images in (a) and (b). Image (c) shows the observed target locations for the image in (a), which were specified by hand.

problem can solved in closed form, so when the misalignment is sufficiently small, this assumption should be made.

3.3 Camera-to-mirror calibration

The camera-to-mirror transformation can be found by minimizing the error between observed projections and predicted reprojections with respect to the camera-to-mirror transformation. Because we assume that the three-dimensional positions of the target points are known, but not the position of the omnidirectional camera relative to those points, we must also include the world-to-camera transformation in the minimization, although the world-to-camera transformation is not subsequently needed and can be discarded once calibration is complete. We use Levenberg-Marquardt to perform the minimization. Here, we briefly describe our error function, and refer the reader to [8] for details of Levenberg-Marquardt, which is widely used.

The world-to-camera transformation is given by the mapping $p_c = R_c p_w + t_c$, where $R_c(\alpha_c, \beta_c, \gamma_c)$ and $t_c = (t_{cx}, t_{cy}, t_{cz})$ are the rotation and translation, respectively, and $\alpha_c, \beta_c, \gamma_c$ are the Euler angles specifying the rotation. So, the world-to-camera transformation is described by a total of six parameters. In contrast, the camera-to-mirror transformation is described by an alternative parameterization, which excludes rotation about the mirror's axis. Specifically, $p_m = R_m(p_c + t_m)$ where $R_m(\beta_m, \gamma_m)$ and $t_m = (t_{mx}, t_{my}, t_{mz})$ are the rotation and translation, for a total of five parameters.

Suppose that $\Pi : \mathcal{R}^3 \mapsto \mathcal{R}^2$ is the projection described in Section 3.2, which gives the image location for a three-dimensional point expressed in the camera coordinate system. It follows that the projection of a point p_i specified in the world system is $\Pi(R_c p_i + t_c)$. Then, if x_i is the observed projection of the known target location p_i, the function we minimize is:

$$\chi^2 = \sum_{i=1}^{n} \parallel x_i - \Pi(R_c p_i + t_c) \parallel^2 \qquad (3)$$

where n is the number of observed targets. Because Π depends on the camera-to-mirror transformation, (3) is minimized with respect to the five camera-to-mirror parameters as well as the six world-to-camera parameters.

3.4 Equiangular omnidirectional cameras

The camera described by Chahl and Srinivasan[1] is a non-single-viewpoint camera that provides approximately uniform angular image resolution. That is, if the camera is placed at the center of a sphere, and points placed along a longitudinal line of the sphere are separated by a uniform angular distance $\Delta\phi$, then the images of these points are separated by some approximately uniform distance along the image's radial lines.

Equiangular cameras are a refinement of this design that provide exact uniformity, as shown in Figure 2(a), and we have designed and fabricated several of these cameras[7]. Figures 2(b) and 2(c) show the effects of misalignment on the projections this design produces. For the experiments described here, we have used a 16 mm lens with a distance of 14 cm between the camera center and mirror apex. The details of the mirror's profile are given in [7].

4. Results

4.1 Overview

To determine whether the calibration algorithm produces the correct camera-to-mirror transformation, and whether

(a)   (b)   (c)
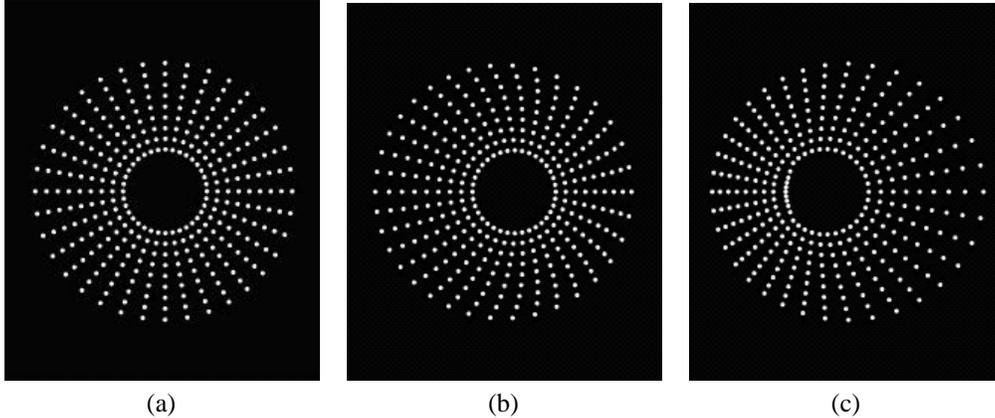
Figure 2. When the equiangular camera's mirror axis and camera axis are identical, lines in the environment parallel to this axis appear as radial lines in the image, as shown in (a). When these axes are not identical, the resulting image is distorted. Images (b) and (c) show the distortion that occurs if the mirror is translated approximately 2 cm along its x axis, or rotated approximately 0.2 radians about its y axis, respectively.

	Calibration A	Calibration B	Calibration C
Sequence 1	1.92	1.91	1.34
Sequence 2	4.01	3.85	1.40
Sequence 3	6.50	6.30	1.36

Table 1. Average reprojection errors for each of the nine calibrations, in pixels.

the full projection model improves the accuracy of shape-from-motion and stereo matching, we have captured three image sequences with increasing amounts of misalignment between the mirror and camera axes. In Sequence 1, the mirror and camera axes are aligned with the same accuracy we might expect from an off-the-shelf omnidirectional camera. In Sequence 2, the camera is significantly rotated about its center of projection. This is both a translation and a rotation of the mirror in the camera coordinate system, so the resulting distortion is a combination of those shown in Figures 2(b) and 2(c). Sequence 3 is similar to Sequence 2, but with a more gross rotation about the camera's center. For each sequence, we have also captured a calibration image. The calibration images for Sequences 1 and 3 are those shown in Figure 1(a) and 1(b); the image locations of the calibration targets for the image in 1(a) are shown in 1(c).

For each of these sequences, we have performed three calibrations. Calibration A assumes that the mirror and camera have precisely the correct relative positions; for the mirror used in these tests, the mirror apex should be 14.0 cm from the camera center. Calibration B assumes that the mirror and camera axes are identical, but that the mirror may be vertically translated with respect to the camera away from

the ideal 14.0 cm location. Calibration C makes no assumptions about the relative positions of the camera and mirror, i.e., all five degrees of freedom are recovered.

4.2 Correctness of estimated parameters

Some basic statistics from these nine calibrations are shown in Tables 1 and 2. Table 1 shows the average reprojection errors, i.e., the average distance in pixels between the observed target image locations and those predicted by the estimated world-to-camera and camera-to-mirror transformations. As one might expect, these show calibration models A and B, which assume that the mirror and camera axes are identical, model the observations more poorly as these two axes are rotated relative to each other, whereas model C, i.e., the full calibration, is a sufficiently strong model in these cases.

Table 2 gives the resulting estimates of the camera-to-mirror transformation, which differ primarily in t_x, and the estimated standard deviations. The standard deviations were computed empirically assuming a 0.25 pixel standard error in the target location observations. Although one might guess that simultaneously estimating the world-to-

	β (radians)	γ (radians)	t_x (cm)	t_y (cm)	t_z (cm)
Sequence 1	-0.0075 ± 0.00087	-0.042 ± 0.0017	0.0052 ± 0.0035	-0.057 ± 0.0055	-14.1 ± 0.014
Sequence 2	0.0015 ± 0.0012	-0.039 ± 0.0023	0.21 ± 0.0053	-0.041 ± 0.0075	-14.2 ± 0.028
Sequence 3	0.0069 ± 0.0013	-0.040 ± 0.0018	0.37 ± 0.0056	-0.054 ± 0.0056	-14.2 ± 0.028

Table 2. Estimates and standard deviations for the five parameters of the full calibration.

	Predicted	Observed	Error Distance
Sequence 1	(318.7, 248.8)	(315.5, 245.0)	5.0 pixels
Sequence 2	(286.1, 246.1)	(280.5, 244.0)	6.0 pixels
Sequence 3	(260.7, 248.2)	(258.5, 244.0)	4.7 pixels

Table 3. The predicted and observed image locations of the mirror apex, in pixels, for the full calibration of each sequence.

camera and camera-to-mirror transformations from a single image would produce estimates that are sensitive to observation errors, these standard deviations, which are insignificant compared to the range of possible parameter values, show that this is not the case.

One independent check on the physical correctness of the full calibration can be performed by comparing the observed image location of the mirror apex to the reprojection of the apex predicted by the camera-to-mirror transformation. These reprojections are shown in Table 3, along with the observed locations. We have taken the observed location to be the image location of the center of the mirror screw, which attaches the mirror to the rig and passes through the mirror axis. In each case the predicted and observed centers are quite close.

4.3 Shape-from-motion accuracy

Shape-from-motion methods recover the motion of a camera and the three-dimensional positions of points in the camera's environment from the two-dimensional image locations of the points, i.e., from the projections. We have implemented a shape-from-motion method for omnidirectional cameras that is similar to Szeliski and Kang's widely known method for conventional cameras[11]. This method uses Levenberg-Marquardt to minimize the difference between the observed projections and those predicted by the camera and point estimates, and produces optimal estimates assuming uniform, isotropic gaussian observation errors.

To help determine whether the camera-to-mirror transformation recovered by the calibration is physically accurate, we have computed the accuracy of the shape-from-motion algorithm for each of our nine pairs of sequence and calibration model. Each of the three sequences consists forty-one images, captured by moving the camera by

hand in a "U" shape on the calibration lab's optical bench. Therefore, the images are similar to the calibration images shown in Figure 1, but show the scene from a large range of views. The "U" motion was chosen to ensure some camera motion parallel to each of the walls and therefore improve the estimation accuracy for points on all walls.

The results are summarized in Table 4. For each of the nine tests, the average reprojection error is shown before the slash, and an average range error is shown for each test after the slash. The range errors are the average error in distance to each of the calibration targets, as measured from the middle camera position in the forty-one-image sequence. The results closely follow the pattern of Table 1: range is recovered accurately in Sequence 1 for all three calibration models, and in all sequences for the full calibration model.

4.4 Epipolar curve accuracy

Stereo ranging applications exploit the known relative position between two cameras to constrain the search for a point from one image to a single curve in the second image. More specifically, if L is the three-dimensional ray that results from backprojecting the image of the point in the first image, then the curve to search in the second image is just the reprojection of L in the second image. For the familiar case of perspective cameras, these reprojections are the epipolar lines. For non-single-viewpoint omnidirectional cameras, the curve has no analytic form, but can be approximated by performing the backprojection and reprojection explicitly.

If the position between the two cameras and their projection model are known to sufficient accuracy, the search can be performed only along the curve, rather than, e.g., in a band of several pixels on either side of the curve. This is a great computational saving and is required for real-time

	Calibration A	Calibration B	Calibration C
Sequence 1	0.40 pixels / 3.3 cm	0.41 pixels / 3.4 cm	0.37 pixels / 2.0 cm
Sequence 2	1.1 pixels / 9.9 cm	1.1 pixels / 9.6 cm	0.43 pixels / 1.9 cm
Sequence 3	1.9 pixels / 15.9 cm	1.92 pixels / 15.2 cm	0.38 pixels / 1.9 cm

Table 4. Average shape-from-motion reprojection error and point range error for each sequence. The reprojection errors are given in pixels; the range errors in centimeters.

	Calibration A	Calibration B	Calibration C
Sequence 1	0.68 pixels	0.69 pixels	0.64 pixels
Sequence 2	1.3 pixels	1.4 pixels	0.71 pixels
Sequence 3	2.1 pixels	2.1 pixels	0.64 pixels

Table 5. The average distance from observed points to the corresponding epipolar curve.

stereo applications. So, the accuracy of epipolar curves is another important measure of appropriateness for each calibration model.

Table 5 shows this accuracy for each of the three sequences and three calibration models used in our shape-from-motion experiments. Each entry is an average over ordered pairs of images within the sequence and over observed features. The pattern is the same as in Tables 1 and 4: all three calibrations are sufficiently accurate for the first sequence, and the full calibration is sufficiently accurate for all three sequences.

5. Conclusions

We have presented an imaging model for omnidirectional cameras that accounts for the full rotation and translation between the camera and mirror, and a method for recovering this relative position from observations of known points in a single image. The method is general in that any combination of camera and mirror can be calibrated, including non-single-viewpoint combinations. For single viewpoint cameras, where the advantages of a single viewpoint can be exploited only if the camera and mirror are assumed to be aligned, this algorithm can be used to verify the alignment accuracy.

We have presented a sequence of experimental results using sequences with increasing amounts of misalignment between the camera and mirror. These results confirm that the method produces an accurate estimate of the relative position between the camera and mirror, that the estimate is not sensitive to small observation errors, and that one image is sufficient to recover the transformation. Additional experiments show that shape-from-motion and stereo matching can be performed accurately if the full model is used, even if the camera and mirror are severely misaligned.

References

[1] J. Chahl and M. Srinivasan. Reflective surfaces for panoramic imaging. *Applied Optics*, 36(31):8275–8285, 1997.

[2] C. Geyer and K. Daniilidis. Catadioptric camera calibration. In *Proceedings of the Seventh International Conference on Computer Vision*, pages 398–404. IEEE, September 1999.

[3] J. Gluckman and S. K. Nayar. Ego-motion and omnidirectional cameras. In *IEEE International Conference on Computer Vision*, pages 999–1005, Bombay, 1998.

[4] S. B. Kang. Catadioptric self-calibration. In *IEEE Computer Vision and Pattern Recognition*, pages 201–207, Hilton Head Island, South Carolina, 2000.

[5] Intel Corporation. Open source computer vision library. http://www.intel.com/research/mrl/research/opencv/.

[6] S. K. Nayar. Catadioptric omnidirectional camera. In *IEEE Computer Vision and Pattern Recognition*, pages 482–488, 1997.

[7] M. Ollis, H. Herman, and S. Singh. Analysis and design of panoramic stereo vision using equi-angular pixel cameras. Technical Report CMU-RI-TR-99-04, Carnegie Mellon University Robotics Institute, Pittsburgh, PA, 1999.

[8] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, Cambridge, 1992.

[9] D. Strelow, J. Mishler, S. Singh, and H. Herman. Extending shape-from-motion to noncentral omnidirectional cameras. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Wailea, Hawaii, October 2001.

[10] T. Svoboda, T. Pajdla, and V. Hlavec. Epipolar geometry for panoramic cameras. In *Fifth European Conference on Computer Vision*, pages 218–232, Freiburg, Germany, June 1998.

[11] R. Szeliski and S. B. Kang. Recovering 3D shape and motion from image streams using nonlinear least squares. *Journal of Visual Communication and Image Representation*, 5(1):10–28, March 1994.