Catadioptric Stereo For Robot Localization

Adam Bickett
CSE 252C Project
University of California, San Diego

Abstract

Stereo rigs are indispensable in real world 3D localization and reconstruction, yet they are costly and the additional camera adds complexity. Catadioptric systems offer an inexpensive alternative, with the bonus of being taken from the same camera. I explore the design and calibration of such a rig, and begin to evaluate its utility for robot localization.

1. Introduction

Catadioptric vision systems enable the collection of rich stereo image information using only a single camera and mirrors. Gluckman and Nayar have investigated the theory behind catadioptric stereo, from the simple, two planar mirror case [5], to rectified stereo with both planar and non-planar sets of mirrors [6],[7]. Gluckman and Nayar cite the benefits of catadioptric stereo, including an additional constraint on the fundamental matrix F relating the cameras, and lack of inter-camera variation.

One important application of stereo camera systems is in robot navigation. Typical robot navigation systems involve dense data collection using costly multi-camera rigs, with the goal of ongoing localization and map-building (SLAMB) [10]. Other works, such as [11], and [1], have used catadioptric systems to gain omnidirectional or panoramic robot vision. I was unable to find the use of catadioptric systems in a simple, sparse data application. One of the main benefits of catadioptric rigs is their relative simplicity and economy, and this is an area that seems largely unexplored. The end goal of this work is to study the utility of an inexpensive catadioptric stereo rig for aiding in the localization of a robot whose main navigation is through odometry and laser sensors.

2. Design

The first decision in the design of a catadioptric system is the type of mirror used. Because the goal of this project is to minimize the complexity of the system while maintaining low cost, my design incorporates inexpensive planar mirrors. Multiple planar catadioptric rig configurations
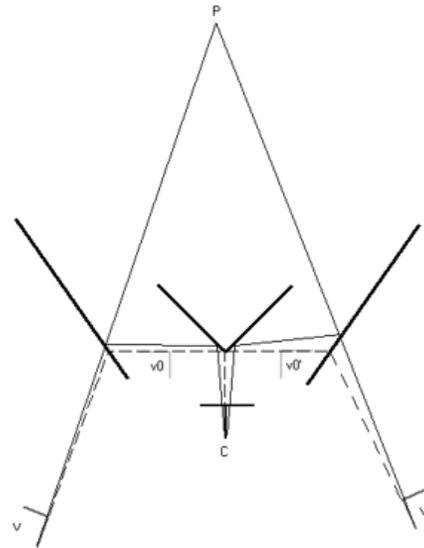

Figure 1: The catadioptric rig setup.

have been proposed in the literature, using from 1 to 5 mirrors [7], [4]. Many of these configurations place the virtual cameras in awkward orientations with respect to the real camera. My design uses four mirrors: a perpendicular mirror pair facing the camera, which splits the image, and two angled mirrors, which create virtual cameras with a high degree of overlap and oriented in roughly the same direction as the physical camera, and is shown in Figure 1. This is actually not a novel design, as I had originally thought. Similar designs have been proposed in the literature in both [9] and [4].

The location of these two virtual cameras can be seen in the mirrors in Figure 2(b).

In order to generate a useful stereo pair, the objective is to maximize the overlapping field of view between the two imaging virtual cameras (shown as v and v' in Figure 1), while minimizing the required rotation. This requires the angles of the mirrors to be set such that the views are slightly "cross-eyed". Unfortunately, this implies that the images captured by the camera will not be rectified, because there is necessarily a rotation between the virtual

Figure 2: (a)The rig. (b) View of virtual cameras

cameras.

The relation between the virtual cameras is captured by a series of reflections about each mirror, as is explained in [5] for the two-mirror case. A reflection transform can be defined by

$$D = \begin{bmatrix} I - nn^T & 2dn \\ 0 & 1 \end{bmatrix}$$

Where n is the normal to the mirror surface, and d is the distance between the mirror and the camera optical center. In our case, the location of intermediate virtual cameras v_0 and v_0' is given by reflections D_1 and D_1', respectively, where D_1 is the reflection about the splitting mirror. Our imaging virtual camera v can be defined by an additional reflection D_2 about the angled mirror, yielding:

$$v = D_2 D_1 c$$

Then the extrinsic orientation of the virtual cameras is given by inverting the reflections back to c and then applying the appropriate reflections:

$$D = D_2 D_1 D_{1'}^{-1} D_{2'}^{-1}$$

,

and because reflection transforms are their own inverse, this is simply

$$D = D_2 D_1 D_{1'} D_{2'}$$

.

It was pointed out in [5] that for the two mirror case, the extrinsic translation of the rig will be limited to the plane defined by mirror normals n and vn', and the axis of the extrinsic rotation will be orthogonal to the plane, having the axis $n \times n'$. In the general four mirror case these limitations do not hold, because it is not guaranteed that the four mirror surface normals will be co-planar. In practice, however, with the mirrors mounted on a flat surface, the virtual cameras will be limited to approximately planar motion.

In keeping with a frugal approach, the camera I selected for the rig is a VGA-resolution webcam. The split in the mirrors takes up approximately 40 pixels of the image, so the resulting stereo pair consists of 480x300 images. After rectification, the usable image size is somewhat smaller. While it is theoretically possible, using the above transformations, to precisely measure for a certain baseline and rotation between the virtual cameras, it is very difficult to adjust for this with sufficient accuracy. It is more practical to adjust the mirrors to obtain the desired image overlap and view, and calculate the extrinsics in the calibration process. The current mirror orientations were adjusted for best image pairs of scenes in the range between 1 m and 5 m.

3   Calibration

The researchers in [7] cite that one benefit of catadioptric systems is the relative ease of setup in comparison to traditional stereo camera rigs. Because there is only one camera, they argue that the internal calibration parameters should be identical between the two views. In addition, in some cases there is a planar motion constraint (described in the previous section) on the extrinsic relationship of the virtual cameras that removes a degree of freedom from the fundamental matrix relating the cameras. Furthermore, synchronization between the cameras is clearly not an issue. My attempts at calibration revealed that catadioptric rigs pose their own problems.

For the purposes of this project, I used the Matlab Calibration Toolbox [3], and partially implemented automatic calibration using calibration components available in Intel's OpenCV [8], which provided the main software framework for this project. The calibration process has 3 main goals:

removing distortion, finding the camera's intrinsic parameters, and finding the camera's extrinsic parameters.

3.1 Intrinsic Parameters

The intrinsic parameters of the camera can be expressed as

$$\mathbf{K} = \begin{bmatrix} fs_x & fs_\theta & o_x \\ 0 & fs_y & o_y \\ 0 & 0 & 1 \end{bmatrix}$$

With f being the camera focal length, and s_x and s_y giving the pixel aspect. Because it is difficult to extract f from the pixel aspect terms, the product terms f_x and f_y, which combine the scaling/skew and the focal length, are found instead. The optical center, or principal point, of the camera is defined by o_x and o_y. s_θ is the pixel skew, but is assumed in our case to be 0, i.e. each pixel is an axis-aligned rectangle.

Despite what one may expect, the intrinsic parameters of the two virtual cameras in the catadioptric setup are not identical. This can be seen by looking at the principal point, or camera center, defined as (o_x, o_y) above. As can be seen in Figure 1, the splitting of the camera view leaves each virtual camera with an image plane on only one side of the principal point. Thus the principal point is not on the visible portion of the image plane for either camera, but, in this rig, is in fact discarded in the splitting of the images. The minimization in the calibration process does find this, but because the calibration of the cameras is done individually, the center point is not consistent (it is usually off by about 10-15 pixels). This could harm the quality of the rectification, as well as interfere with disparity-depth measurements.

The focal length was generally well determined among the calibration efforts and consistent when quality image pairs were used.

3.2 Extrinsic Parameters

The extrinsic parameters can be given as

$$\mathbf{g} = \begin{bmatrix} R & \mathbf{T} \\ \mathbf{0}^\top & 1 \end{bmatrix}$$

Placing the left camera at the world origin, we only need solve for the rotation and translation $\mathbf{g}'$ that describe the pose of the right camera with respect to the origin. Unlike the distortion coefficients and the principal point, calibration reliably was able to determine the extrinsic relationship between the cameras. The final rig orientation was characterized extrinsically by $\mathbf{T} = (-132.08036, 0.53307, 13.68733)$, with rotation vector $\phi = (-0.02220, 0.39865, -0.04318)$. Note that this system does exhibit near planar motion. This also gives us the baseline between the cameras, which is $\|T\| = 13.28$ cm.
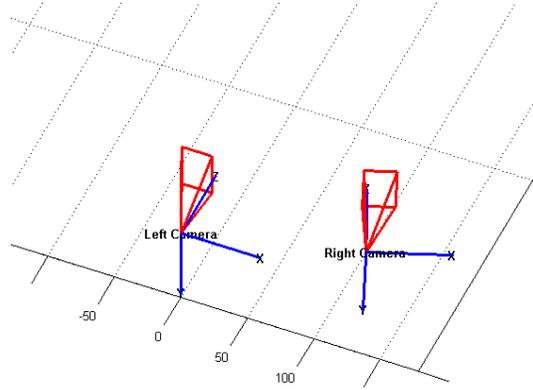

Figure 3: A view of the virtual camera poses

3.3 Distortion Correction

Calibration typically found very small distortion coefficients, likely due to the small lens of the webcam used. The commonly employed radial distortion model alters a projected point $\mathbf{x}$ as

$$\mathbf{x}' = (1 + k_1 r^2 + k_2 r^4 + k_3 r^6)\mathbf{x}$$

where r is the distance from the image center. The only consistently significant distortion term was the fourth order radial distortion term, k_2. Curiously, this value for one camera view was often estimated to be of opposite sign in the other camera view. Furthermore, the high variance of this value led me to suspect that this term was being used to compensate for noise (such as imprecise checkerboard corner locations) or other unmodeled distortion sources, and did not represent a reliable value. My final approach was to calculate the distortion coefficients of the webcam independently of the mirror setup, and remove the small amount of distortion before splitting the images. The idea behind this approach is that additional noise introduced by the mirrors is unlikely to be well-described by the parameterization of the radial or tangential distortion. In the end, the effects of undistorting the image were small.

3.4 Practical Issues

Other issues with catadioptric rigs complicate the calibration procedure. Some of these issues are discussed by Bailey et al. in [4], in which the non-parametric calibration of a very similar rig is investigated, the motivation being that the mirrors introduce unmodelable distortions. Bailey's approach involved creating a per-pixel distortion map. After spending countless hours in the lab fiddling with getting consistent calibration results just right, a calibration model that doesn't presuppose an approximating model is attractive.

The main difficulties I came across in the calibration was wide variability in the calibration measurements. A few factors of the catadioptric rig setup exacerbated these issues. The first issue is capturing quality calibration images. I used a planar checkerboard pattern across multiple views, which clearly must project to the visible image plane of both virtual camera views. With the current baseline and rig setup, this limited the calibration rig to being at least 1 m from the cameras. At this range, and with the camera resolution, this limits the precision of the location of the checkerboard corners, which in turn hurts the precision of the calibration minimization. Also an issue is a blurring and loss of image quality near the outside of the edges in each view, due to the high angle of incidence of the incoming light, which does not always reflect reliably off imprecise mirrors. In order to achieve a quality calibration, I took a large quantity of calibration images, and through iterative trial and error, pruned them to locate the images for which good corners could be extracted for each pair. Selecting sets of good calibration images in this manner helped find reasonable calibration parameters with lower variation.

4 Evaluation

After achieving proper calibration parameters, the goal of testing the utility of the catadioptric system remained. The first step here is to calculate the rectifying transform that results in the two views being related by a translation along the X axis, with the epipoles mapped to $\inf$. This results in the desirable property of having the epipolar lines, along which matches are made between the images, lie on the horizontal scanlines. I used the approach from [12] to obtain this transformation.

4.1 Depth Accuracy

With rectified images, calculating the depth of a corresponding point in the two images is simple, given by

$$Z = \frac{fB}{d}$$

where d is the pixel disparity between the images, f is the focal length of the camera in mm (790 mm for this rig), and B is the baseline between the cameras (13.28 cm). In our case, this gives the following approximate pixel resolutions per depth.

I used the framework I developed in OpenCV for the upcoming extension of this project to robot navigation to find feature matches, and then check the disparity between feature pairs Currently this uses simple Forstner corner detection with normalized cross-correlation to determine correspondences. The detected objects were soda cans, placed at

Table 1: Depth Accuracy

Distance	Approx. Pixel Width	Measured Error
1 m	1 cm	5.3 cm
2 m	4 cm	11.9 cm
3 m	8 cm	16.2 cm
4 m	15 cm	41.3 cm
5 m	20 cm	74.7 cm
10m	1m	–

distances measured from the approximate left virtual camera location. Table 1 displays the error averaged over five different views. This simply gives a rough idea of performance, as it is limited by the placement of the feature matching with NCC, the detection of the interest points, and the subsequent testing on random objects around the lab did find higher errors, likely due to the quality of the correspondences.

4.2 Depth Map

The above approach did not use the convenience of the horizontal epipolar lines in its finding of correspondences. Depth maps, in contrast, use this property of rectified image pairs to attempt to determine a depth for each pixel of the image. I created depth maps using the Birchfield dynamic programing algorithm (from [2]) in OpenCV. A sample created image is shown in Figure 5. Significant depth information is clearly recovered.

Figure 4: A depth map and a corresponding view

5 Conclusion

Catadioptric systems offer an alternative to expensive stereo rigs, while at the same time eliminating the issues inherent in comparing images from two different cameras. Calibration of catadioptric systems (at least with this rig design) is made more difficult by low quality mirrors and the relationship between the views. But these are relatively small practical issues, and given a well-calibrated and stable rig

setup, a catadioptric system should be a viable low-cost alternative to traditional stereo rigs. I plan on testing the use of this design in the future to aid robot navigation.

Acknowledgments

Tom Duerig provided significant help in the design and building of the catadioptric rig. I also used some portions of his SMORs code to aid in the feature matching code to be used in the upcoming robot.

References

[1] R. Benosman, E. Deforas, and J. Devars. A new catadioptric sensor for the panoramic vision of mobile robots. In *OMNIVIS '00: Proceedings of the IEEE Workshop on Omnidirectional Vision*, page 112, Washington, DC, USA, 2000. IEEE Computer Society.

[2] Stan Birchfield and Carlo Tomasi. Depth discontinuities by pixel-to-pixel stereo. In *ICCV*, pages 1073–1080, 1998.

[3] Jean-Yves Bouget. Camera calibration toolbox for matlab.

[4] G. Gupta D. Bailey, J. Seal. Non-parametric calibration for catadioptric cameras. Image and Vision Computing - New Zealand, 2005.

[5] Joshua Gluckman and Shree K. Nayar. Planar catadioptric stereo: Geometry and calibration. *cvpr*, 01:1022, 1999.

[6] Joshua Gluckman and Shree K. Nayar. Catadioptric stereo using planar mirrors. *Int. J. Comput. Vision*, 44(1):65–79, 2001.

[7] Joshua Gluckman and Shree K. Nayar. Rectified catadioptric stereo sensors. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(2):224–236, 2002.

[8] Intel. Opencv.

[9] H. Mathieu and F. Devernay. Systeme de miroirs pour la stereoscopie.

[10] Stephen Se, David G. Lowe, and James J. Little. Vision-based mobile robot localization and mapping using scale-invariant features. In *ICRA*, pages 2051–2058, 2001.

[11] Niall Winters, Jose Gaspar, Gerard Lacey, and Jose Santos-Victor. Omni-directional vision for robot navigation. *omnivis*, 00:21, 2000.

[12] Jana Kosecka Yi Ma, Stefano Soatto and Shankar Sastry. *An Invitation to 3-D Vision: From Images to Geometric Models*. Springer-Verlag, New York, 2003.