Single camera stereo using planar parallel plate

Chunyu Gao and Narendra Ahuja
Beckman Institute, University of Illinois at Urbana-Champaign
Email: cgao@uiuc.edu, ahuja@vision.ai.uiuc.edu

Abstract

A system of using a planar parallel plate to achieve single camera stereo has been proposed by Nishimoto and Shirai[5].Their work was based on an assumption that the optical axis of the camera was equally displaced by a tilted planar plate with two fixed tilt angles. Such assumption is invalid in general. In this paper, we propose a general framework to more accurately model such a single camera system using a planar parallel plate and to deal with arbitrary orientation of the plate. Our model has no limitation on the FOV of the camera. The proposed framework includes: a mathematical formulation to calculate the displacement of scene points; a plate calibration method to determine the plate intrinsic parameters as well as its extrinsic pose, and a generalized correspondence method which is capable of dealing with arbitrary number of images captured from different plate poses; Experimental result is included to validate the proposed framework.

.

1. Introduction

Stereopsis has been one of the most widely explored topics of 3D vision. A classical stereo system typically includes two cameras placed side by side to capture stereo image pairs. The depth information of the captured scene is calculated from the disparity map of these two images. Instead of using two or more cameras, one can also sequentially capture image pairs by repositioning a single camera. The advantage of using a single camera over the two-camera systems is that the identical intensity response of the stereo pairs captured with the same camera can improve the accuracy of correspondence matching.

Besides the approach of displacing a single camera, several other single-camera systems have also been investigated. Each of them has its own unique features. Adelson and Wang designed a plenoptic camera system which uses a lenticular array placed in front of the image plane to control the light path, and the depth information is extracted by analyzing a set of sub-images [1]. Lee and Kweon presented a bi-prism stereo camera system which forms the stereo pairs on the left and right halves of a single CCD sensor by a bi-prism [3]. Nene and Nayar investigated stereo imaging using different types of mirrors including planar, ellipsoidal, hyperboloidal and paraboloidal mirror [4].

The work most closely related to this paper is that of Nishimoto and Shirai [5]. A planar parallel plate is placed in front of the camera as shown in the Fig.1. By changing the tilt angle of the plate about vertical axis which is normal to the optical axis of the camera, the image of scene object shifts horizontally. In their experiment, two images corresponding to two fixed poses of the plate are captured. Feature points are extracted by looking for zero-crossings on images filtered by LOG (Laplacian of Gaussian) filter. The correspondences of these feature points are found by searching the closest feature points on the horizontal scan line. To allow such searching, the plate has to be carefully calibrated with the camera, the two tilt angles of the plate have to be large.

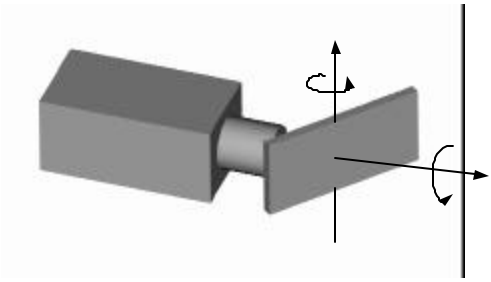

Fig. 1. Concept illustration (the plate can be rotated about either a horizontal axis or a vertical axis)

In this paper, we propose a general framework to more accurately model such a single camera system and to allow arbitrary orientation of the plate. Our model has no limitation on the FOV of the camera. Furthermore, we not only achieve two-view stereo, but also multi-baseline stereo by rotating the plate around the optical axis or by changing the tilt angle of the plate, as illustrated in the Fig. 1. With the large number of extra images captured at multiple plate poses, it is possible to construct depth maps with higher accuracy.

2. Mathematical model

2.1 Planar parallel plate

It is known from optics that light ray passing through a planar plate will encounter a lateral displacement ***d*** given by:

$$d = t \sin I \left[1 - \sqrt{\frac{1 - \sin^2 I}{n^2 - \sin^2 I}} \right] \qquad (1)$$

Where ***I*** is incident angle; ***n*** is refractive index and ***t*** is thickness of the plate [7]. In this paper, the refraction index ***n*** and thickness ***t*** are referred to as the intrinsic parameters of the plate.

2.2 Extrinsic parameters and essential points

Like most vision systems, the camera is assumed to be a pinhole which can be simplified as a projection center along with it's intrinsic parameters (focal length ***f***, image center ***(u_0,v_0)*** and distortion factor ***k***). A parallel plate is placed arbitrarily in front of the camera. Besides the intrinsic parameters (***n, t***), extrinsic parameters described below are also necessary to characterize the plate. For the proposed system, the plate position is not important because the lateral displacement of a ray depends only on its incident angle. On the other hand, the plate pose is critical for calculating the displacement. Instead of using a normal vector to denote the pose of a plate, we introduce the concept of an essential point to specify the

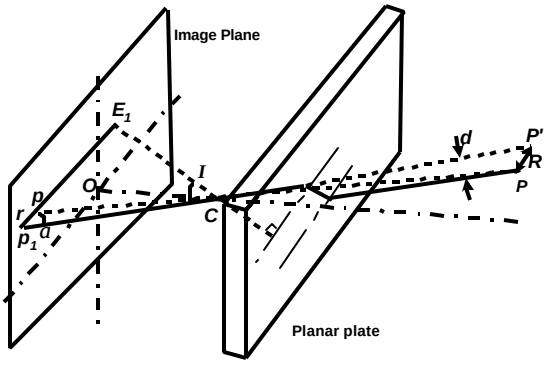

Fig 2. Projection of a 3D point through a plate

pose of a plate. As illustrated in Fig. 2, for each plate pose, there exists a unique point that is the intersection of the plate normal, passing through the camera projection center, with the camera image plane. We call this point as essential point. The essential point of the plate under pose ***i*** is denoted as ***$E_i(u_e, v_e, f)$***. Given the focal length of the camera, two coordinates ***(u_e, v_e)*** of the essential point are enough to specify a plate pose and are referred to as extrinsic parameters of the plate. With the two intrinsic parameters, a total of four parameters fully characterize the plate optical properties and its pose.

Similar to the use of an epipole in epipolar geometry, the use of essential point limits the search for correspondences to 1D. In Fig 2, an object point ***P(x,y,z)*** is imaged to point ***$p_1(u_1, v_1)$*** on the sensor plane through the projection center ***C*** and the plate. We denote ***p(u,v)*** to be the image point of the point ***P*** through the projection center ***C*** without the plate. ***E_1*** is the essential point of the plate under a certain pose (pose 1). It is evident that the essential point ***E_1***, image point ***p*** and ***p_1*** are collinear. We define the line passing through those three points as essential line of the image point ***p*** (or ***p_1***; or 3D point ***P***) corresponding to plate pose 1. If the plate is repositioned to a new pose (pose 2) with the essential point ***E_2***, then the image of the 3D point ***P*** through the plate can be found on the new essential line ***E_2p***. Therefore, like the classical two-camera stereo case, the correspondence search is limited to 1D; in other words, search can be performed on the essential lines. However, unlike an epipolar line pair for which points on the lines can be

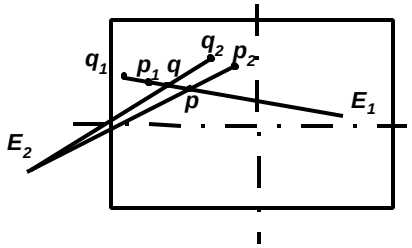

Fig 3. Properties of essential line pairs

matched one by one in order, there are only two points matched on an essential line pair. As illustrated in Fig 3, the match point of the neighbor pixel ***q_1*** of pixel ***p_1*** cannot be found on the essential line ***E_2p***, but on the essential line ***E_2q***. Furthermore, any essential line of the 3D point ***P*** passes through its image point ***p***. These properties set up the foundation for the correspondence matching method we will discuss later and enable us to perform only 1D search for correspondence.

2.3 Estimation of the depth of the objects

For a given object point, we assume that the point is shifted by the plate in a plane parallel with image plane. As shown in Fig 2, the 3D point ***P*** is shifted to point ***P′*** by the plate and the displacement is denoted as ***R***. The image points of ***P*** and ***P′*** through the projection center ***C*** are ***p*** and ***p_1*** respectively. The displacement of ***P*** in image space is the distance between ***p*** and ***p_1*** and is denoted as ***r***. If we know both ***R*** and ***r***, then the depth of ***P*** can be calculated by simple triangulation.

$$z = f * R / r \qquad (2)$$

The problem of finding the displacement ***r*** is referred to as correspondence problem, and will be discussed in

section 3.2. The displacement $\boldsymbol{R}$ can be calculated by equation (1) and following equations:

$$\cos a = \frac{\overline{p_1C} \bullet \overline{p_1E_1}}{\left|\overline{p_1C}\right| \bullet \left|\overline{p_1E_1}\right|} \quad (3)$$

$$R = d / \sin a \quad (4)$$

$$\cos I = \frac{\overline{p_1C} \bullet \overline{E_1C}}{\left|\overline{p_1C}\right| \bullet \left|\overline{E_1C}\right|} \quad (5)$$

where: $\boldsymbol{a}$ is the angle between line $\boldsymbol{p_1E_1}$ and $\boldsymbol{p_1C}$; $\boldsymbol{d}$ is the lateral displacement of the chief ray of $\boldsymbol{P}$ which is given by equation(1); $\boldsymbol{I}$ is the angle between line $\boldsymbol{p_1C}$ and line $\boldsymbol{E_1C}$ which is the input of equation(1).

3. Plate calibration and correspondence matching.

3.1 Plate calibration

Calibration is a critical issue for many systems. Good calibration can compensate for systematic errors and improve system accuracy. For the proposed system, the goal of calibration is to accurately estimate both the intrinsic and extrinsic parameters of a plate defined in the previous section. We also anticipate that the calibration process will minimize the error resulting from the aberrations introduced by the plate.

For conciseness, we rewrite Equations (1~5) as a single function:

$$z = F(X, r, X_C, P) \quad (6)$$

where: $\boldsymbol{z}$ is the depth of a given 3D point; $\boldsymbol{X}$ denotes the plate parameters ($\boldsymbol{u_e, v_e, n, t}$); $\boldsymbol{r}$ is the displacement in image space for a given point; $\boldsymbol{X_c}$ denotes the camera intrinsic parameters; $\boldsymbol{P}$ is a 3D object point.

For plate calibration, we use a set of grid points with known depth and estimate the plate parameters by solving equation (6) using nonlinear least square solver. The displacement in image space is calculated by comparing corner coordinates. The calibration procedures is as follows:

(1) Perform standard camera calibration to estimate the intrinsic parameters of the camera [8];
(2) Fix the camera on an optical table and place a calibration pattern in front of the camera at a certain distance.
(3) Capture a reference image of the calibration pattern without the plate, extract corner features, estimate the pose of the calibration pattern relative to the camera, and determine the depth **z** of these grid points on the pattern.
(4) Insert the plate in front of the camera at a series of desired poses, capture a set of images of the calibration pattern corresponding to each pose, and extract corners in the images.
(5) Estimate the displacement $\boldsymbol{r}$ in the image space for each corner feature by comparing two sets of corners extracted from the reference image and an image with the plate at a given pose.
(6) Given $\boldsymbol{r}$, $\boldsymbol{z}$ and the coordinates of the corners, determine the intrinsic and extrinsic parameters by solving nonlinear equation (6) using nonlinear least square solver.

Two sets of calibration results corresponding to two plate poses are listed in table 1. We can see that the root-mean-squared error is less than 0.1 pixels. This indicates that the proposed plate parametric model is very accurate and is capable of fully characterizing the plate and its poses. The actual thickness of the plate we used is 12.96mm. For pose 1, the calibrated thickness is almost identical to the ground truth. But for the 50-deg tilt angle (pose 2), the discrepancy in the thickness is about 0.9mm, and in the index is about 0.1. We observe that such discrepancies at a large tilt angle might relate to two aspects. Firstly, the plate, bought from a glass shop, is not optical-grade glass. Its refractive index and thickness may not be constant across the plate. Secondly, the aberrations introduced by the plate at a large tilt angle will cause additional displacement in the image space. In order to obtain accurate plate intrinsic parameters (thickness and index), it would help to perform calibration at a small tilt angle, and then fix intrinsic parameters when calibrating the poses at large tilt angles. However, keeping the intrinsic parameters as variables at large tilt angles could possibly better compensate for the artifacts caused by the aberrations and the imperfect glass quality than varying just two extrinsic parameters. In our implementation, we treat all four parameters as variables for all the plate poses.

Table 1. Two sets of calibration results

	Pose1 (0° tilt)	Pose2 (50° tilt)
u_e (pixel)	733.45	3289.9
v_e (pixel)	484.6	465.1
Refractive Index	1.46	1.576
Thickness (mm)	12.974	13.84
Residual (pixel)	0.049	0.075

3.2 Correspondence

There are two ways to extract depth from the images captured by varying the plate pose. The straightforward method is to directly conduct 2D search in the images to identify the best matching features or pixels, and then to use the recovered disparity map to calculate the depth of the scene by equations (1) through (5). Obviously, 2D search will be expensive. Based on the analysis in section 2.2, we could search for the best matching points on the essential lines. In order to do so, besides the images we captured at different plate poses, we introduce

a reference dummy image, which can be thought of as the image captured by the camera without the plate, to bridge those images.

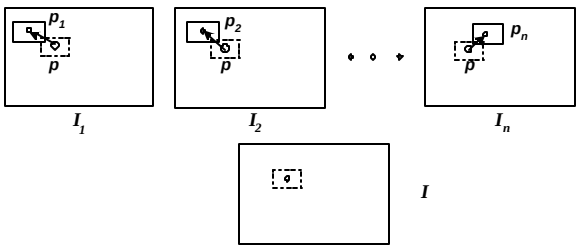

Fig.4. Illustration of the multi-image matching

Fig. 4 shows a set of images $I_1, I_2, \ldots, I_n$ which correspond to n plate poses. The image on the second row is the dummy image I. Given a pixel p in image I, one can calculate the displacement of corresponding 3D point P by equations 3. If the depth of the point P is known, then corresponding pixels $p_1, p_2, \ldots, p_n$ in the images $I_1, I_2, \ldots, I_n$ can be computed. Normally, $p_1, p_2, \ldots, p_n$ are not integers and the intensity values of these pixels have to be interpolated from their neighbors. However, the depth of the given 3D point P is something we want to find. So instead of using one depth, we specify a depth range $Zmin$ to $Zmax$ and for each depth level, identify the corresponding pixels in given images. The depth of given point P will be the one for which the intensity values of $p_1, p_2, \ldots, p_n$ best match. Of course, matching one pixel will not give us a good depth estimate in general; a block around pixel p has to be used. The matching cost function could be any of the functions commonly used in standard stereo case. In our implementation, we use SAD (sum of absolute differences)[2,6]. Note that one advantage of such a single viewpoint camera is that the image intensity of corresponding points can be safely assumed to be invariant regardless of scene and surface complexities.

4. Experimental results and conclusion

In this section, we show that the proposed method is able to accurately estimate depth map. In our experiment, a Sony DFX900 color camera with 1280x960 resolution was used to capture images. The plate was mounted on a rotation stage to change the plate pose about vertical axis. Totally 15 poses were calibrated using the calibration method described in the Section 3.1. Figure 5 shows a box (Fig. 5a) and its front surfaces reconstructed from two images with block size 9x9 (Fig. 5b) and from multiple images (15 images totally) with block size 3x3 (Fig. 5c). While two views are sufficient to recover the depth, multiple views yield smoother depth estimates. More quantitive and complete analysis of the accuracy of depth estimation will be carried out in future work. In summary, the proposed framework is capable of extracting depth with a single camera and is able to handle arbitrary number of images captured from different plate poses. By increasing the number of images, our method has the potential of recovering depth map and 3D shapes with high accuracy.

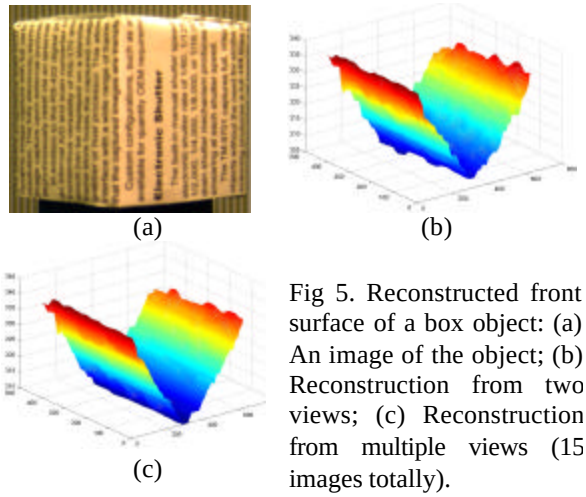

Fig 5. Reconstructed front surface of a box object: (a) An image of the object; (b) Reconstruction from two views; (c) Reconstruction from multiple views (15 images totally).

Acknowledgements

The support of National Science Foundation under grant ECS 02-25523 is gratefully acknowledged

References

[1] Edward H. Adelson and John Y.A. Wang, "Single Lens Stereo with a Plenoptic Camera", *In IEEE Transactions on Pattern Analysis and Machine Intelligence*, VOL. 14, NO. 2., February 1992.

[2] U. R. Dhond and J. K. Aggarwal. "Structure from stereo – a review", *IEEE Transaction on systems, Man, and Cybern.*, 19(6): 1489-1510, 1989

[3] Doo Hyun Lee and In So Kweon, "A novel stereo camera system by a biprism", *IEEE Transactions on Robotics and Automation*, 16(5), Oct. 2000

[4] S. Nene and S. Nayar. Stereo with mirrors. *In Proceedings of the 6th International Conference on Computer Vision*, Bombay, India, January 1998

[5] Y. Nishimoto, and Y, Shirai,. "A feature-based stereo model using small disparities", *Proc. of the IEEE International Workshop on Industrial Applications of Machine Vision and Machine Intelligence*. Seiken Symposium, 192-196, Tokyo 2-5 Feb. 1987

[6] Daniel Scharstein and Richard Szeliski, "A taxonomy and evaluation of dense two frame stereo correspondence algorithm", *International Journal of Computer Vision*, 47:7-42, April 2002

[7] Warren J. Smith "Modern optical Engineering", McGraw-Hill, 3rd edition, NY, 2000

[8] Z., Zhang, "A flexible new technique of camera calibration", IEEE Transactions on Pattern Analysis and Machine Intelligence, *22(11): 1330-1334, 2000.*