MVA'90 IAPR Workshop on Machine Vision Applications Nov. 28-30,1990, Tokyo

WHEN SHOULD WE CONSIDER LENS DISTORTION IN CAMERA CALIBRATION

Yi-Ping Hung, Sheng-Wen Shieh

Institute of Information Science
Academia Sinica
Nankang, Taipei, Taiwan
Republic of China

ABSTRACT

In this paper, calibrating a camera refers to the process of determining the transformation between a 3D object coordinate system and the 2D image coordinate system of the camera. Knowing this transformation, 3D position of a feature point seen in two cameras can be determined by using triangulation. Suppose a distortion-free lens is had, this 3D-2D transformation can be well approximated by a linear model. Unfortunately, most off-the-shelf lenses have a certain amount of distortion. If our goal is to get highly accurate 3D measurements, we have to consider lens distortion in camera calibration. However, considering lens distortion will not only complicate the camera calibration procedure, but also complicate the subsequent on-line processing such as feature-point correspondence and camera re-calibration. It is hence desirable to avoid considering lens distortion whenever the distortion-induced error is tolerable. This work investigates the effect of neglecting lens distortion, and proposes a simple efficient calibration algorithm which can either exclude or include lens distortion -- depending on whether one can neglect lens distortion and still meet the accuracy requirement.

I. INTRODUCTION

To infer 3D objects using two or more images, it is essential to know the relationship between the 2D image coordinate system and the 3D object coordinate system. This relationship can be described by the following two transformations:

(i) Perspective projection of a 3D object point onto a 2D image point -- Given an estimate of a 3D object point and its error covariance, we can predict its projection (mean and covariance) on the 2D image. This is useful for reducing the searching space in matching features between two images, or for hypothesis verification in scene analysis.

(ii) Back projection of a 2D image point to a 3D ray -- Given a 2D image point, there is a ray in the 3D space that the corresponding 3D object point must lie on. If we have two (or more) views available, an estimate of the 3D point location can be obtained by using triangulation. This is useful for inferring 3D information from 2D image features.

The above 3D-2D relationship can be specified by a column vector $\boldsymbol{\beta}$, which contains the geometric camera parameters specifying camera orientation and position, focal length, lens distortion, optical axis misalignment, and pixel size. Determining this 3D-2D relationship, or equivalently, estimating $\boldsymbol{\beta}$, is called (geometric) camera calibration.

The techniques for camera calibration can be classified into two categories: one that considers lens distortion [2] [8] [9], and one that neglects lens distortion [3] [6] [7]. A typical linear technique that does not consider lens distortion is the one estimating the perspective transformation matrix $\mathbf{H}$ [7]. The estimated $\mathbf{H}$ can be used *directly* for forward and backward 3D-2D projection. If necessary, given the estimated $\mathbf{H}$, the geometric camera parameters $\boldsymbol{\beta}$ can be easily determined [4][5][6].

Faig's method [2] is a good representative for those considering lens distortion. For methods of this type, equations are established that relate the camera parameters to the 3D object coordinates and 2D image coordinates of the calibration points. Nonlinear optimization techniques is then used to search for camera parameters with an objective to minimize residual errors of those equations. One disadvantage of this kind of method is that a good initial guess is required to start the nonlinear search.

A few years ago, Tsai proposed an efficient two-stage technique using the "radial alignment constraint" [8]. His method involves a direct solution for most of the calibration parameters and some iterative solution for the other parameters. Some drawbacks of Tsai's method are pointed out in [9]. Our experimental results in section IV also show that Tsai's method can be worse than the simple linear method of [7] if lens distortion is relatively small.

Recently, [9] shows some experimental results using a two-step method. The first step involves a closed-form solution based on a distortion-free camera model, and the second step improves the camera parameters estimated in the first step by taking into account lens distortion. This method overcomes the initial guess problem in the nonlinear optimization, and is more accurate than Tsai's method according to our experiments.

However, considering lens distortion will not only complicate the camera calibration procedure, but also complicate the subsequent on-line processing such as feature-point correspondence (in stereo) and camera re-calibration (in the case of having a moving camera). Notice that epipolar line is no longer a straight line if lens distortion is taken into account. Moreover, when lens distortion is small, if the noise in the 2D feature extraction is relatively large or the number of the calibration points is relatively small, the calibration results based on distortion camera model can be worse than those based on linear camera model. The question is then, "when should we consider lens distortion in camera calibration?" or "when does it worth all the trouble to consider lens distortion?" This work represents an initial effort toward the answer of the question.

II. CAMERA MODEL

Consider the pinhole camera model with lens distortion, as shown in Figure 1. Let P be an object point in the 3D space, and $\mathbf{r}_O = (x\ y\ z)^t$ be its coordinates, in inches, with respect to a fixed object coordinate system (OCS). Let the camera coordinate system (CCS), also in inches, have its x-y plane lying on the front image plane (such that x axis is parallel with the horizontal direction of the image, and y axis is parallel with the vertical one), and its z axis aligned with the optical axis of the lens (see Figure 1). Let $\mathbf{r}_C = (x_C\ y_C\ z_C)^t$ be the coordinates of the 3D point P with respect to the CCS. Suppose there is no lens distortion, the corresponding image point of P on the front image plane would be Q (see Figure 1). However, due to the effect of lens distortion, the actual image point is Q'. Let $\mathbf{s}_I = (u\ v)^t$ denote the 2D coordinates (in pixels), with respect to the computer image coordinate system (ICS), of the actual image point Q'.

As shown in Figure 2, the 3D-2D transformation from $\mathbf{r}_O$ to $\mathbf{s}_I$ can be divided into the following four steps:

1. Translation and rotation from the OCS to the CCS

The transformation from $\mathbf{r}_O$ to $\mathbf{r}_C$ can be expressed as

$$\tilde{\mathbf{r}}_C = \mathbf{T}_C^O \tilde{\mathbf{r}}_O \qquad \text{with} \qquad \mathbf{T}_C^O = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0\ 0\ 0 & 1 \end{bmatrix} \tag{1}$$

where tilde ($\tilde{\ }$) denotes homogeneous coordinates [1], $\mathbf{t} = (t_1\ t_2\ t_3)^t$ is a translation vector, and $\mathbf{R}$ is a 3x3 rotation matrix determined by the three Euler angles, ϕ, θ, ψ, rotating about the z, y, z axes sequentially.

2. Perspective projection from a 3D object point in the CCS to a 2D image point on the front image plane

Let f be the "effective focal length", and let $\mathbf{s}_F = (u_F\ v_F)^t$ be the 2D coordinates (in inches) of the undistorted image point Q lying on the front image plane, i.e., the x-y plane of the CCS. Then, we have

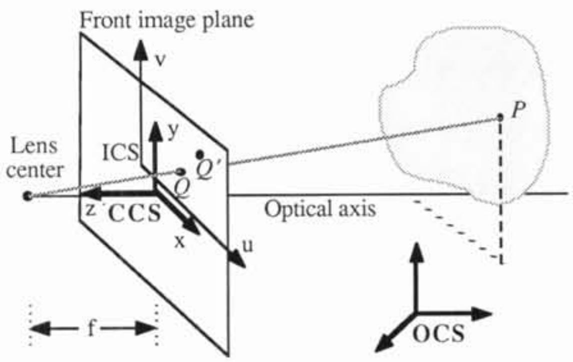

Figure 1. Pinhole camera model with lens distortion.
OCS -- object coordinate system (3D)
CCS -- camera coordinate system (3D)
ICS -- computer image coordinate system (2D)

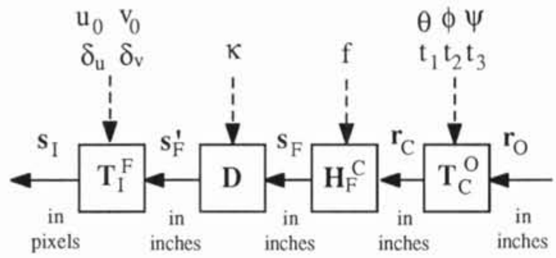

Figure 2. Relation between different transformation matrices.

$$\tilde{\mathbf{s}}_F = \mathbf{H}_F^C \tilde{\mathbf{r}}_C \qquad \text{with} \qquad \mathbf{H}_F^C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1/f & 1 \end{bmatrix} \tag{2}$$

3. Lens distortion from Q to Q'

For practical reasons, we consider only the first term of the radial lens distortion, i.e.,

$$\mathbf{s}'_F = (1 + \kappa \|\mathbf{s}_F\|^2)\, \mathbf{s}_F \tag{3}$$

where $\mathbf{s}'_F = (u'_F\ v'_F)^t$. In this paper, κ has the unit of inch^{-2}.

4. Scaling and translation of 2D image coordinates

The transformation from $\mathbf{s}'_F$ (in inches) to $\mathbf{s}_I$ (in pixels) involves (i) scaling from inches to pixels, and (ii) translation due to misalignment of the sensor array with the optical axis of the lens. Hence,

$$\tilde{\mathbf{s}}_I = \mathbf{T}_I^F \tilde{\mathbf{s}}_F \qquad \text{with} \qquad \mathbf{T}_I^F = \begin{bmatrix} 1/\delta_u & 0 & u_0 \\ 0 & 1/\delta_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{4}$$

where δ_u and δ_v are the horizontal and vertical pixel spacing (inches/pixel), u_0 and v_0 are the coordinates (in pixels) of the origin of the CCS in the computer image coordinate system.

Using the above notations for camera parameters, $\boldsymbol{\beta} = (t_1\ t_2\ t_3\ \phi\ \theta\ \psi\ f\ \kappa\ \delta_u\ u_0\ v_0)^t$. The vertical scaling factor δ_v is not included here because it is a known parameter when we use a solid state camera -- otherwise, only the ratios f/δ_u and f/δ_v can be determined.

Notice that, suppose there is no optical distortion (i.e., $\kappa = 0$), the relationship between $\mathbf{r}_O$ and $\mathbf{s}_I$ can be expressed as a linear transformation:

$$\tilde{\mathbf{s}}_I = \mathbf{H}\tilde{\mathbf{r}}_O \qquad \text{i.e.,} \qquad \begin{bmatrix} u\cdot w \\ v\cdot w \\ w \end{bmatrix} = \begin{bmatrix} H_{11} & H_{12} & H_{13} & H_{14} \\ H_{21} & H_{22} & H_{23} & H_{24} \\ H_{31} & H_{32} & H_{33} & H_{34} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{5}$$

where $\mathbf{H} = \mathbf{T}_I^F \mathbf{H}_F^C \mathbf{T}_C^O$ and D is identity operator (see Figure 2).

III. CAMERA CALIBRATION TECHNIQUES

First, consider the linear calibration method based on distortion-free camera model. The problem is to estimate $\boldsymbol{\beta}$, but keep $\kappa = 0$, given a set of 3D calibration points $\mathbf{r}_k = (x_k\ y_k\ z_k)^t$ and their corresponding 2D image coordinates $\mathbf{s}_k = (u_k\ v_k)^t$, $k = 1, \ldots, N$. Instead of estimating $\boldsymbol{\beta}$ directly, we first estimate the composite parameters $\{H_{ij}\}$ in Eq. (5)

using Sutherland's method [7], and then decompose $\{H_{ij}\}$ into β, uniquely, with a geometric procedure similar to Strat's method [6] (due to the space limitation, please refer to chapter 2 of [5] for details.) As long as the amount of distortion is reasonably small, the β estimated from the above linear method is accurate enough for many applications. If the resulted β is not satisfactory, it can be used as an initial estimate in the following nonlinear calibration procedure.

The nonlinear calibration procedure searches for the estimate of β by minimizing an objective function $J(\beta)$ with respect to β. In this paper, $J(\beta)$ is chosen to be

$$\sum_{k=1}^{N} \| s_F^{(3)}(\mathbf{r}_k;\beta_1) - s_F^{(2)}(\mathbf{s}_k;\beta_2) \|^2 \qquad (6)$$

where $\beta_1 = (t_1\ t_2\ t_3\ \phi\ \theta\ \psi\ f)^t$ and $\beta_2 = (\kappa\ \delta_u\ u_0\ v_0)^t$;

$s_F^{(3)}(\mathbf{r}_k;\beta_1)$ is the projection of the 3D point $\mathbf{r}_k$ on the front image plane, and can be obtained by using $\tilde{s}_F^{(3)} = \mathbf{H}_F^C\, \mathbf{T}_C^O\, \tilde{\mathbf{r}}_k$;

$s_F^{(2)}(\mathbf{s}_k;\beta_2)$ is the corrected (undistorted) image point on the front image plane corresponding to the 2D image observation $\mathbf{s}_k$, and can be obtained by using

$s_F^{(2)} \approx (1 - \kappa \|s'_F\|^2)\, s'_F$ where $\tilde{s}'_F = (\mathbf{T}_I^F)^{-1}\, \tilde{s}_k$.

Other objective functions can also be chosen. However, the $J(\beta)$ chosen in Eq. (6) is consistent with the error measure used for testing calibration accuracy in stereo vision application. Also, when minimizing (6) with quasi-Newton method, the iterative estimates are found to converge quite well.

When the amount of distortion is large, the initial estimate for β (with $\kappa = 0$) obtained from the linear calibration procedure can be far away from the true value, which will make the subsequent iterative estimation less likely to converge correctly. By noticing that the distortion is much more severe near the image boundary than that in the central image part, we can use only the calibration points observed in the central part of the image to perform the initial linear calibration, while use all the calibration points for the subsequent nonlinear minimization step. This technique is used for obtaining $\hat{\beta}_J$ shown in the next section.

IV. EXPERIMENTAL RESULTS

To evaluate the accuracy of the camera calibration for 3D vision application, it is necessary to define certain kind of error measure. The measure adopted in this paper is the *3D angular error*, i.e, the angle $\measuredangle POP'$ where P is the 3D test point, O is the lens center, and OP' is the 3D ray back projected from the observed 2D image of P. A 3D angular error of 0.015° is roughly equivalent to "1 part in 4000" (because $\tan(0.015°) \approx 1/4000$).

Let $\hat{\beta}_H$ represent the value of β estimated by using the linear calibration method described in the last section. Let $\hat{\beta}_J$ represent the estimate for β obtained by minimizing $J(\beta)$ in (6) with $\hat{\beta}_H$ as the initial estimate. Let $\hat{\beta}_T$ be the estimate obtained by Tsai's method presented in [8]. Figures 3 - 6 compare the 3D angular errors associated with $\hat{\beta}_H$, $\hat{\beta}_J$, and $\hat{\beta}_T$, i.e., the estimates for β obtained with three different calibration methods -- the first one is based on distortion-free model while the other two take into account radial lens distortion. The size of the images used in the experiments is 256x242 pixels.

Figures 3 - 5 are obtained with synthetic data, so that the true camera parameters are known exactly. Each data point in these Figures is the average value from ten random trials. In these simulation, we assume the 3D positions of the calibration points are known exactly, and the only source of measurement noise is the error in estimating the image coordinates of the calibration points, i.e., the 2D observation noise. The reason for doing so in the simulation is because, for our applications, the 3D measurement noise is easier to be controlled such that it has much smaller effect than the 2D observation noise has. Let σ denote the standard deviation of the 2D observation noise. For Figure 3, σ is chosen to be 0.1 pixel, and the true κ is -0.3 inch^{-2} which corresponds to roughly 4 to 5 pixels distortion near the four image corners. However, the distortion at the central part of the image is much smaller than one pixel because the distortion is a cubic function of the radius. The purpose of Figure 3 is to show how 3D angular error decreases as the number of calibration points N increases. Figure 4 shows that as the 2D observation noise increases, the 3D angular error also increases, i.e., the camera calibration accuracy decreases, for all the three calibration methods.

A few interesting results are shown in Figure 5. **(i)** As the amount of distortion (or $|\kappa|$) gets larger, the accuracy of $\hat{\beta}_H$ decays rapidly; however, the accuracy of $\hat{\beta}_J$ does not vary much. The inaccuracy of $\hat{\beta}_H$ at large $|\kappa|$ is due to the incapability of modeling distortion. **(ii)** When the amount of distortion is relatively small, say, $|\kappa| < 0.25$, $\hat{\beta}_T$ is worse than $\hat{\beta}_H$. This must have been owing to the instability caused by the extra freedom of estimating κ. **(iii)** The accuracy of $\hat{\beta}_J$ is

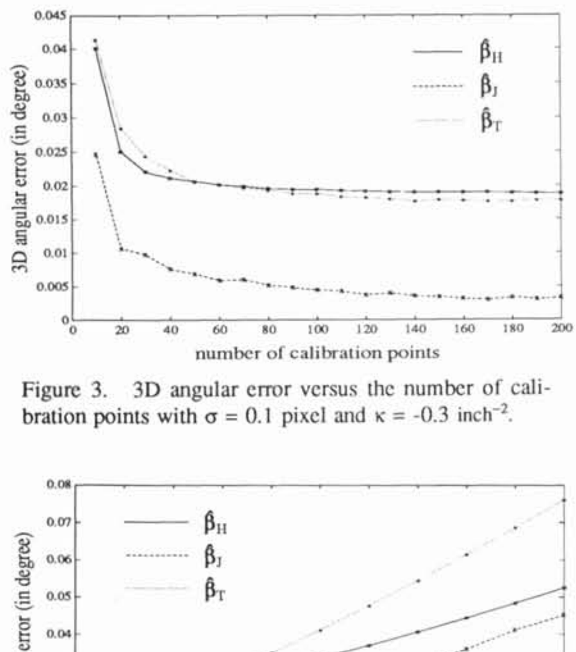

Figure 3. 3D angular error versus the number of calibration points with σ = 0.1 pixel and κ = -0.3 inch^{-2}.

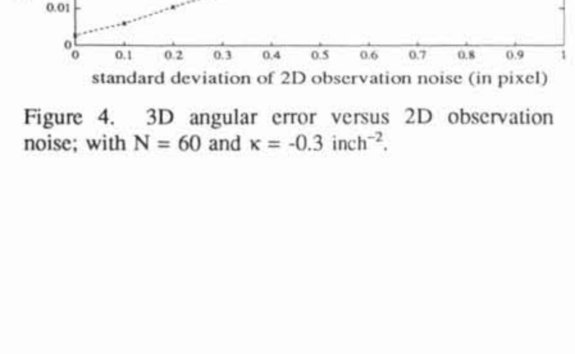

Figure 4. 3D angular error versus 2D observation noise; with N = 60 and κ = -0.3 inch^{-2}.

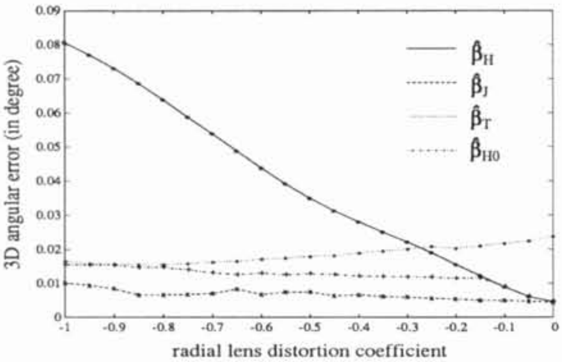

Figure 5. 3D angular error versus lens distortion κ; with $N = 60$ and $\sigma = 0.1$ pixel.

almost always better than that of $\hat{\beta}_H$ (although some exceptions do exist in our extensive simulation). This is quite natural since we use $\hat{\beta}_H$ as the initial estimate for the nonlinear minimization of $J(\beta)$. **(iv)** Let $\hat{\beta}_{H0}$ be the estimate obtained by using the linear method and using only the calibration points located in the central image area. The radius of the central image area, R_0, is chosen such that the distortion incurred in the central area is always smaller than 1 pixel. From Figure 5, We can see that the accuracy of $\hat{\beta}_{H0}$ is much better than that of $\hat{\beta}_H$ for large $|\kappa|$, and is even better than that of $\hat{\beta}_T$ in general. (Of course, the accuracy measure of $\hat{\beta}_H$ is obtained by testing only on the test points located within the central area.)

Figure 6 shows the experimental results using real images taken by a SONY XC-39 CCD camera at 19 different positions. The 3D control points used are the centers of 61 circular dots located on two 3D planes. It is found that the average 3D angular error using the linear calibration method is 0.0149°, which is "1 part in 4000". This 3D angular error roughly corresponds to 2D image error of 0.15 pixels. If we consider lens distortion and include the nonlinear minimization step, the average 3D angular error can be reduced to 0.0122°. This improvement is much smaller than that expected from the simulation. This is mainly because that, in real experiments, there is always noise in estimating the 2D image coordinates of test points. These 2D observation noise of test points will make the computed inaccuracy measure of $\hat{\beta}_J$ appear to be much larger than it really is. Another reason for the improvement being smaller is probably because the distortion model is not exact.

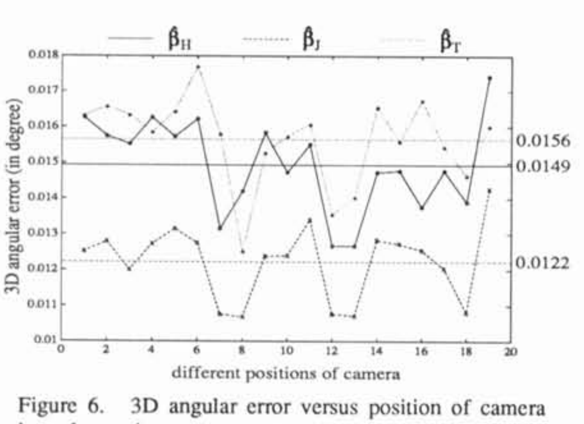

Figure 6. 3D angular error versus position of camera in real experiment.

V. CONCLUDING REMARKS

From the above experiments, we make the following observations:

(i) To increase the calibration accuracy to a certain extent, we can either increase the number of the calibration points or decrease the 2D observation noise.

(ii) For small lens distortion, $\hat{\beta}_H$ is usually good enough for 3D applications. Unless we can increase the accuracy of 2D feature extraction accordingly, it may not worth the efforts to increase the calibration accuracy by the nonlinear minimization step.

(iii) With the modified linear method which uses only the central calibration points, we can enjoy the linearity property of the camera model without losing too much the calibration accuracy. The major sacrifice is that we do not use the 2D features outside the central area for 3D inference. (For $|\kappa| = 0.3$, the radius of the central area is about 105 pixels for a 256 by 242 image.) This small drawback is no problem at all for a 3D active vision system.

(iv) Surprisingly, Tsai's method is not as good as expected. This is partly because it does not estimate the image center (at the beginning), and partly because it does not use all the information contained in the calibration points.

REFERENCES

[1] R.O. Duda, P.E. Hart, *Pattern Recognition and Scene Analysis*, Wiley, New York, 1973.

[2] W. Faig, ''Calibration of Close-Range Photogrammetry Systems: Mathematical Formulation,'' Photogrammetric Engineering and Remote Sensing, Vol. 41, No. 12, 1975, pp. 1479-1486.

[3] O.D. Faugeras, G. Toscani, ''The Calibration Problem for Stereo,'' Proceedings Conf. on Computer Vision and Pattern Recognition, 1986, pp. 15-20.

[4] S. Ganapaphy, ''Decomposition of Transformation Matrices for Robot Vision,'' Proceedings Int. Conf. on Robotics and Automation, 1984, pp. 130-139.

[5] Y.P. Hung, ''Three Dimensional Surface Reconstruction Using a Moving Camera A Model-Based Probabilistic Approach,'' Ph.D dissertation, Division of Engineering, Brown University, Providence, R.I.; also, Technical Report LEMS-63, 1989.

[6] T.M. Strat, ''Recovering the Camera Parameters from a Transformation Matrix,'' DARPA Image Understanding Workshop, 1984.

[7] I. Sutherland, ''Three-Dimensional Data Input by Tablet,'' Proceedings of the IEEE, Vol. 62, No. 4, 1974, pp 453-461.

[8] R.Y. Tsai, ''A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses,'' IEEE Journal of Robotics and Automation, Vol. RA-3, No. 4, 1987, pp. 323-344.

[9] J. Weng, P. Cohen, M. Herniou, ''Calibration of Stereo Cameras Using a Non-linear Distortion Model,'' Proc. 10th Inter. Conf. on Pattern Recognition, 1990, pp. 246-253.