行政院國家科學委員會補助專題研究計畫成果報告

※※※※※※※※※※※※※※※※※※※※※※※※※
※　　　　　　　　　　　　　　　　　　　　　　　※
※機械臂視覺伺服控制系統中之主動式攝影機參數校準※
※　　　　　　　　　　　　　　　　　　　　　　　※
※※※※※※※※※※※※※※※※※※※※※※※※※

計畫類別：☑個別型計畫　　□整合型計畫

計畫編號：NSC89－2213－E009－130

執行期間：88 年 8 月 1 日至 89 年 7 月 31 日

計畫主持人：林錫寬 教授

本成果報告包括以下應繳交之附件：

□赴國外出差或研習心得報告一份

□赴大陸地區出差或研習心得報告一份

□出席國際學術會議心得報告及發表之論文各一份

□國際合作研究計畫國外研究報告書一份

執行單位：

中 華 民 國 89 年 10 月 12 日

行政院國家科學委員會專題研究計畫成果報告

機械臂視覺伺服控制系統中之主動式攝影機參數校準
An Alternative Approach to the Camera Self-Calibration of an Active Vision System

計畫編號: 89N210 NSC89-2213-E009-130

執行期限: 88年08月01日至89年07月31日

主持人:林錫寬教授　交通大學電機與控制系

1 摘要

本計畫提出一個二段式攝影機校正法。第一階段中，藉由直線投影關係提出不考慮透鏡扭曲的新校正法，並在第二階段修正透鏡扭曲係數。我們的方法有利於歸避過多透鏡扭曲係數時引起的不收斂問題，實驗亦有正確的校正結果。

Abstract

This report presents a two-stage camera calibration method. In the first stage, a new calibration method for cameras without lens distortion is used, which is based on the projective relation of a straight line. The distortion coefficients are corrected in the second stage. Our method is useful to circumvent the not convergent problem when consider all types of distortion together. The experiment generates an accurate calibration.

2 Introduction

Camera calibration is important for the application of a vision system to reconstructing 3-D world information from 2-D images. Many techniques have been developed for camera calibration. The perspective transformation matrix method proposed by Faugeras and Toscani [1] and the method of the focus of expansion (FOE) presented by Ma [2] are two typical methods of the intrinsic calibration for the case of no lens distortion. The disadvantage of the FOE method is the difficulty to find the accurate FOE because of the digital image quantization. Several methods considering lens distortion incorporates the perspective transformation matrix method. The explicit calibration method of Wei and Ma [3] uses the cross ratio technique to find the distortion center and the radial distortion coefficient.

This report presents a two-stage camera calibration method. The first stage solves the intrinsic and extrinsic parameters for a lens with known distortion coefficients, while the second corrects the distortion coefficients. Our new method is established by using the relationship between a 3-D line and its projective image line. In the second stage, we suggest to take into account different types of distortion in sequence.

3 Calibration Method

3.1 First Stage: No Distortion

We first consider a pinhole camera model without lens distortion. Let E_c and E_i denote the camera frame and the image pixel frame, respectively. The *effective focal length* is denoted by f. The intersection point of the optical axis and the image plane is (u_0, v_0) in pixels.

Let P_i be an 3-D point with the coordinates (x_{ci}, y_{ci}, z_{ci}) with respect to frame E_c. The ideal image coordinates of the projection of the point P_i on the image plane without lens distortion are $(\bar{u}_i, \bar{v}_i)$:

$$\begin{cases} \bar{u}_i - u_0 = \frac{f}{\delta_u}\frac{x_{ci}}{z_{ci}} = f_x\frac{x_{ci}}{z_{ci}}, \\ \bar{v}_i - v_0 = \frac{f}{\delta_v}\frac{y_{ci}}{z_{ci}} = f_y\frac{y_{ci}}{z_{ci}} \end{cases} \tag{1}$$

where δ_u and δ_v are the horizontal and vertical spacing of the image sensor array (mm/pixel). Then, u_0, v_0, f_x, and f_y are the *intrinsic parameters* of a camera.

Let $\mathbf{d}_i$ be a directional vector of the straight line L_i. It is known that the projection images of points on a straight line on the image plane still form a straight line for a lens without distortion. Let the ideal image pixel coordinates of the projection of the points P_i and P_j be $(\bar{u}_i, \bar{v}_i)$ and $(\bar{u}_j, \bar{v}_j)$, respectively. The slope of the image line of the projection of the line L_i is then μ_i/ν_i, where $\mu_i = \bar{u}_j - \bar{u}_i$ and $\nu_i = \bar{v}_j - \bar{v}_i$.

The coordinate transformation from E_w to E_c is a rotation (denoted by $\mathbf{R}$) and the distance from the origin of E_w to E_c is a translation $\mathbf{t}$. It is well known that $\mathbf{p}_i^{<c>} = \mathbf{R}\mathbf{p}_i^{<w>} + \mathbf{t}^{<c>}$, where the superscript "$< \cdot >$" denotes the representation of a vector with respect to a specified frame.

Let r_{ij} denotes the (i, j)th entry of $\mathbf{R}$. We obtain the linear regression form of

$$\mathbf{As} = \mathbf{b} \tag{2}$$

where $\mathbf{A} = [\mathbf{a}_1, \ \cdots, \ \mathbf{a}_m]^T$, $\mathbf{b} = [b_1, \ \cdots, \ b_m]^T$, in which $b_i = -\beta_i d_{i3}^{<w>}$ with $\beta_i = \mu_i\bar{v}_i - \nu_i\bar{u}_i$, and

$$\mathbf{a}_i = \begin{bmatrix} \nu_i d_{i1}^{<w>} \\ \nu_i d_{i2}^{<w>} \\ \nu_i d_{i3}^{<w>} \\ -\mu_i d_{i1}^{<w>} \\ -\mu_i d_{i2}^{<w>} \\ -\mu_i d_{i3}^{<w>} \\ \beta_i d_{i1}^{<w>} \\ \beta_i d_{i2}^{<w>} \end{bmatrix}, \ \mathbf{s} = \frac{1}{r_{33}} \begin{bmatrix} r_{11}f_x + r_{31}u_0 \\ r_{12}f_x + r_{32}u_0 \\ r_{13}f_x + r_{33}u_0 \\ r_{21}f_y + r_{31}v_0 \\ r_{22}f_y + r_{32}v_0 \\ r_{23}f_y + r_{33}v_0 \\ r_{31} \\ r_{32} \end{bmatrix} \tag{3}$$

provided $r_{33} \neq 0$. Each line L_i is given by two points (x_{ci}, y_{ci}, z_{ci}) and (x_{cj}, y_{cj}, z_{cj}).The directional vector $\mathbf{d}_i^{<w>}$ is the displacement of the platform and is equal to the difference of the coordinates of P_i and P_j with respect to frame E_w. In (2), the ideal image $(\bar{u}_i, \bar{v}_i)$ of the point P_i is used to form $\mathbf{a}_i$.

Given m lines L_i, $m \geq 8$, then $\mathbf{s}$ is the solution to the least squares (LS) problem: $\min_{\mathbf{s}} \|\mathbf{As} - \mathbf{b}\|$, where $\mathbf{A} = [\mathbf{a}_1, \ \cdots, \ \mathbf{a}_m]^T$ and $\mathbf{b} = [b_1, \ \cdots, \ b_m]^T$. As $\mathbf{s}$ is obtained, then the orthogonality of the matrix $\mathbf{R}$ allows us to obtain r_{33}, r_{31}, r_{32}, u_0, v_0, f_x, f_y, and matrix $\mathbf{R}$ in order. The sign of r_{33} is reasonable to be known in advance. To make good use of (2), we also assume $|r_{33}| \approx 1$ to keep the components of $\mathbf{s}$ not too large.

3.2 Second Stage: Distortion Coefficients

When a lens has distortion, an observed image can be modeled as an ideal perspective with some distortion terms [4]. Let the real image coordinates are (u_i, v_i). Define $\rho_{ui} \equiv u_i - u_0$, $\rho_{vi} \equiv v_i - v_0$, and $\tilde{f} \equiv f_y/f_x = \delta_u/\delta_v$. Then the distortion model [4] is

$$\begin{aligned} \delta_u(\bar{u}_i - u_0) \ = \ & \delta_u\rho_{ui} + \bar{\kappa}_1\delta_u\rho_{ui}(\rho_{ui}^2\delta_u^2 + \rho_{vi}^2\delta_v^2) \\ & +\bar{\kappa}_2(3\rho_{ui}^2\delta_u^2 + \rho_{vi}^2\delta_v^2) \\ & +2\bar{\kappa}_3\rho_{ui}\rho_{vi}\delta_u\delta_v \\ & +\bar{\kappa}_4(\rho_{ui}^2\delta_u^2 + \rho_{vi}^2\delta_v^2) \end{aligned} \tag{4}$$

$$\begin{aligned} \delta_v(\bar{v}_i - v_0) \ = \ & \delta_v\rho_{vi} + \bar{\kappa}_1\delta_v\rho_{vi}(\rho_{ui}^2\delta_u^2 + \rho_{vi}^2\delta_v^2) \\ & +2\bar{\kappa}_2\rho_{ui}\rho_{vi}\delta_u\delta_v \\ & +\bar{\kappa}_3(\rho_{ui}^2\delta_u^2 + 3\rho_{vi}^2\delta_v^2) \\ & +\bar{\kappa}_5(\rho_{ui}^2\delta_u^2 + \rho_{vi}^2\delta_v^2) \end{aligned} \tag{5}$$

where $\bar{\kappa}_1$ is the radial distortion coefficient, $\bar{\kappa}_2$ and $\bar{\kappa}_3$ are the tangential ones and $\bar{\kappa}_4$ and $\bar{\kappa}_5$ are the prism o However, these coefficients $\bar{\kappa}_i$ are not identifiable for unknown δ_u and δ_v.

We substitute (4) and (5) into (1) to obtain

$$\mathbf{A}_i^*\boldsymbol{\kappa} = \mathbf{b}_i^* \tag{6}$$

where

$$\boldsymbol{\kappa} = \begin{bmatrix} \kappa_1 \\ \kappa_2 \\ \kappa_3 \\ \kappa_4 \\ \kappa_5 \end{bmatrix} \equiv \begin{bmatrix} \bar{\kappa}_1 \delta_u^2 \\ \bar{\kappa}_2 \delta_u \\ \bar{\kappa}_3 \delta_v \\ \bar{\kappa}_4 \delta_u \\ \bar{\kappa}_5 \delta_v \end{bmatrix}, \quad \mathbf{b}_i^* = \begin{bmatrix} f_x \frac{x_{ci}}{z_{ci}} - \rho_{ui} \\ f_y \frac{y_{ci}}{z_{ci}} - \rho_{vi} \end{bmatrix}$$

$$\mathbf{A}_i^* \equiv \left[\begin{matrix} \rho_{ui}(\rho_{ui}^2 + \rho_{vi}^2/\tilde{f}^2) & 3\rho_{ui}^2 + \rho_{vi}^2/\tilde{f}^2 \\ \rho_{vi}(\rho_{ui}^2 + \rho_{vi}^2/\tilde{f}^2) & 2\rho_{ui}\rho_{vi} \end{matrix} \right.$$
$$\left. \begin{matrix} 2\rho_{ui}\rho_{vi} & \rho_{ui}^2 + \rho_{vi}^2/\tilde{f}^2 & 0 \\ \rho_{ui}^2\tilde{f}^2 + 3\rho_{vi}^2 & 0 & \rho_{ui}^2\tilde{f}^2 + \rho_{vi}^2 \end{matrix} \right] \quad (7)$$

Equation (6) provides a LS method to estimate the distortion coefficients $\boldsymbol{\kappa}$, instead of the original ones $\bar{\kappa}_i$. According to our experience, the radial distortion is dominant in the lens distortion, and the prism one is the least important. We then propose a modified iteration way in the following:

Algorithm 1: Camera Calibration.

A1. Let the initial value of $\boldsymbol{\kappa}$ be $\mathbf{0}$.

A2. (First Stage) Let $\kappa_2 = \cdots = \kappa_5 = 0$ and substitute them and the estimated κ_1 into (4) and (5) to compute $\bar{u}_i$ and $\bar{v}_i$. Then solve u_0, v_0, f_x, f_y, $\mathbf{R}$ while considering only the points whose image coordinates are near $(0, 0)$ pixel.

A3. (Second Stage) Assign the parameters u_0, v_0, f_x, f_y, $\mathbf{R}$, and $\boldsymbol{\kappa}$ as those estimated above and reconstruct the 3-D coordinates of the measured points with respect to E_c. Let $\kappa_2 = \cdots = \kappa_5 = 0$ and solve κ_1 by (6) while considering only the points whose image coordinates are far away from $(0, 0)$ pixel.

A4. Repeat *A2* and *A3* until the parameters converge.

A5. Repeat *A2* and *A4*, but let κ_2 and κ_3 vary and still keep $\kappa_4 = \kappa_5 = 0$. If the parameters converge, go to *A6*. Otherwise, stop the algorithm and retain the estimate result at the end of *A4*.

A6. Repeat *A2* and *A3*, but let all $\kappa_2, \ldots, \kappa_5$ vary. If the parameters converge, they are the estimate result. Otherwise, stop the algorithm and retain the estimate result at the end of *A5*. ■

The 3-D space reconstruction method in step *A3* is similar to the stereo vision method, only the distance between two cameras is replaced by the displacement of the camera.

To determine the convergence of the parameters, we compare the residual errors of the LS method, instead of the parameter estimates, between iterations. Both residual errors in stages 1 and 2 are taken into account, since stage 1 uses the central points and stage 2 does the far points.

4 Experiments

The experiment equipment is a three-orthogonal-axis platform with a camera mounted on the z-axis. The 640×480 image has the center at $(0, 0)$ pixel. A small black disk is observed for the subpixel accuracy. The camera are moved along the x-, y-, and z-axis to collect the image points, a virtual $13 \times 13 \times 11$ point-type cuboid will be observed. The distance between the adjacent levels in z- axis or the adjacent points in x- or y-axis is 10 mm. The distance between the lens and the central point of the top level (Level 0) is about 70 mm. We use only eight points for *A2*. The eight points are the conners of the largest rhombus and the ones of the largest rectangle on Level 0 and Level 10, respectively. We select the longest 14 lines connected by these 8 points. The candidate points for *A3* are those on Level 0 and outside of a circle with the radius of 160 pixels, centered at the origin of the image plane. However, we select only 12 evenly radially distributed points for all experiments.

Fig. 1 shows typical experiment results. In Figs. 1(a) and 1(b), the line L_1 is the history of the residual norm of *A4*, while the lines L_1' and L_1'' are those of *A5* and *A6*, respectively. It is apparent that *A6* (i.e., L_1'') is not convergent, so κ_4 and κ_5 should be set to zero. We also did a usual iteration procedure that takes κ_1, κ_2 and κ_3 (and all $\kappa_1, \ldots, \kappa_5$, respectively) into account together in a single iteration loop, the result is depicted as L_2 (and L_3, respectively). L_3 is not convergent like L_1''. It can be seen that stage 2 converges faster than stage 1. However, we ask for the convergence of both stages, so stage 1 determines the required number of iterations. It follows from Fig. 1(a) that the numbers of iterations for L_1 and L_1' are, respectively, 27 and 6, while that for L_2 is 66. According to our experience, the new sequential iteration way mostly requires less iterations and is more efficient.

The average errors and the standard deviations of the 3-D reconstruction coordinates for the final results of *A4* and *A5* is shown in Fig. 1(c) and (d), respectively. The set of estimated parameters additionally comprising the tangential distortion (i.e., L_1') is always better in accuracy than the one with only the radial distortion when the convergent parameters can be obtained.

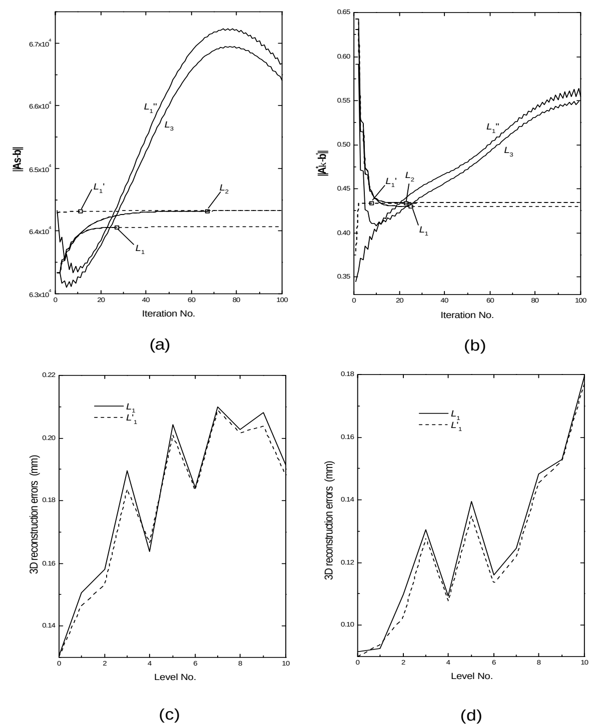

圖 1: Typical experiment results of Algorithm 1: (a) convergent performance of stage 1; (b) convergent performance of stage 2; (c) average 3-D reconstruction errors; (d) standard deviations of 3-D reconstruction errors.

5 Conclusions

This report presents a two-stage camera calibration method. In the first stage, A new calibration method for cameras with no lens distortion is used, which is based on the projective relation of a straight line. The distortion coefficients are corrected in the second stage. Our method suggests three iteration loops, which first considers only the radial distortion, and then encompasses the tangential and the prism distortion in sequence. This is to circumvent the not convergent problem when consider all types of distortion together.

參考文獻

[1] O. D. Faugeras and G. Toscani, "The calibration problem for stereo," in *Conf. on Computer Vision and Pattern Recognition*, pp. 15–19, 1986.

[2] S. D. Ma, "A self-calibration technique for active vision systems," *IEEE Trans. Robotics Automat.*, vol. 12, no. 1, pp. 114–120, 1996.

[3] G. Q. Wei and S. D. Ma, "Implicit and explicit camera calibration: theory and experiments," *IEEE Trans. on Pattern Anal. Machine Intell.*, vol. 16, no. 5, 469–480, 1994.

[4] J. Weng, P. Cohen, and M. Herniou, "Camera calibration with distortion models and accuracy evaluation," *IEEE Trans. on Pattern Anal. Machine Intell.*, vol. 14, no. 10, pp. 965–980, 1992.