Seminar Paper Presentation
A Flexible New Technique for Camera Calibration

Han Xiao hxiao9@jhu.edu Project #4

1. Paper Information

Zhang, Zhengyou. "**A flexible new technique for camera calibration.**" Pattern Analysis and Machine Intelligence, IEEE Transactions on 22.11 (2000): 1330-1334.

2. Paper Introduction

The paper proposes a flexible new technique to easily calibrate a camera. The technique only requires the camera to observe a planar pattern shown at a few (at least two) different orientations. Either the camera or the planar pattern can be freely moved. The motion need not be known. Radial lens distortion is modeled. The proposed procedure consists of a closed-form solution, followed by a nonlinear refinement based on the maximum likelihood criterion. The author provides both computer simulation and real data to test the proposed technique, and very good results have been obtained. Compared with classical techniques that use expensive equipment, the proposed technique is easy to use and flexible. It advances 3D computer vision one step from laboratory environments to real world use.

3. Why Choosing This Paper

Our project is RGBD camera integration into camera augmented mobile c-arm. We have 3 cameras: The X-ray camera, the CCD camera and the Kinect camera. These cameras are mounted near each other; however, their positions and orientations are different. We need to know the relationship between these cameras so that we can co-register images. We need a good camera calibration technique. Currently, there are three main calibration techniques. The first one is photogrammetric calibration. In this method, the calibration is performed by observing a calibration object, whose geometry in 3-D space is known with very good precision; however, this calibration apparatus is very expensive, and we need an elaborate setup. The second method is self-calibration. The calibration is done by moving a camera in a static scene, without a calibration object, and using image information only. This method is not yet mature. There are many parameters to estimate, and it cannot always obtain reliable results. We choose zhang's new camera calibration technique, because it is flexible, robust and low cost, which meets our requirements for the camera calibration for our project.

4. Background Knowledge

A 2D point is

$$\mathbf{m} = [u, v]^T$$

A 3D point is

$$\mathtt{M} = [X, Y, Z]^T$$

For homogeneous coordinate we set

$$\widetilde{\mathbf{m}} = [u, v, 1]^T \; \widetilde{\mathtt{M}} = [X, Y, Z, 1]^T.$$

Then the relationship between a 3D point M and its image projection m is given by

$$s\widetilde{\mathbf{m}} = \mathbf{A}\begin{bmatrix}\mathbf{R} & \mathbf{t}\end{bmatrix}\widetilde{\mathtt{M}}$$

$$\mathbf{A} = \begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$

A is called the intrinsic parameters matrix. R and t are translation and rotation from the camera coordinate system to world coordinate system, which are called the extrinsic parameters

Assume the model plane is on Z = 0 of the world coordinate system. Let's denote the i-th column of the rotation matrix R by ri:

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{A} \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 & \mathbf{t} \end{bmatrix} \begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix}$$

$$= \mathbf{A} \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t} \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}.$$

Therefore, a model point M and its image m is related by a homography H (Up to a scalar):

$$s\widetilde{\mathbf{m}} = \mathbf{H}\widetilde{\mathtt{M}} \quad \text{with} \quad \mathbf{H} = \mathbf{A}\begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t} \end{bmatrix}$$

Set the homography

$$\mathbf{H} = \begin{bmatrix}\mathbf{h}_1 & \mathbf{h}_2 & \mathbf{h}_3\end{bmatrix}$$

$$\begin{bmatrix}\mathbf{h}_1 & \mathbf{h}_2 & \mathbf{h}_3\end{bmatrix} = \lambda \mathbf{A} \begin{bmatrix}\mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t}\end{bmatrix}$$

r1 and r2 are orthonormal for a rotation matrix

$$\mathbf{h}_1^T \mathbf{A}^{-T} \mathbf{A}^{-1} \mathbf{h}_2 = 0$$
$$\mathbf{h}_1^T \mathbf{A}^{-T} \mathbf{A}^{-1} \mathbf{h}_1 = \mathbf{h}_2^T \mathbf{A}^{-T} \mathbf{A}^{-1} \mathbf{h}_2$$

These are the 2 constraints on the intrinsic parameters. The paper also discuss a geometric interpretation of these two constrain using abstract concepts like the image of the absolute conic and circular points.

5. Closed Form Solution

$$\mathbf{B} = \mathbf{A}^{-T}\mathbf{A}^{-1} \equiv \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{12} & B_{22} & B_{23} \\ B_{13} & B_{23} & B_{33} \end{bmatrix}$$
$$= \begin{bmatrix} \frac{1}{\alpha^2} & -\frac{\gamma}{\alpha^2\beta} & \frac{v_0\gamma - u_0\beta}{\alpha^2\beta} \\ -\frac{\gamma}{\alpha^2\beta} & \frac{\gamma^2}{\alpha^2\beta^2}+\frac{1}{\beta^2} & -\frac{\gamma(v_0\gamma-u_0\beta)}{\alpha^2\beta^2}-\frac{v_0}{\beta^2} \\ \frac{v_0\gamma-u_0\beta}{\alpha^2\beta} & -\frac{\gamma(v_0\gamma-u_0\beta)}{\alpha^2\beta^2}-\frac{v_0}{\beta^2} & \frac{(v_0\gamma-u_0\beta)^2}{\alpha^2\beta^2}+\frac{v_0^2}{\beta^2}+1 \end{bmatrix}$$

Note that B is symmetric, defined by a 6D vector

$$\mathbf{b} = [B_{11}, B_{12}, B_{22}, B_{13}, B_{23}, B_{33}]^T$$

Set

$$\mathbf{h}_i = [h_{i1}, h_{i2}, h_{i3}]^T$$

$$\mathbf{h}_i^T \mathbf{B} \mathbf{h}_j = \mathbf{v}_{ij}^T \mathbf{b}$$

$$\mathbf{v}_{ij} = [h_{i1}h_{j1}, h_{i1}h_{j2} + h_{i2}h_{j1}, h_{i2}h_{j2},$$
$$h_{i3}h_{j1} + h_{i1}h_{j3}, h_{i3}h_{j2} + h_{i2}h_{j3}, h_{i3}h_{j3}]^T$$

Then

$$\begin{bmatrix} \mathbf{v}_{12}^T \\ (\mathbf{v}_{11} - \mathbf{v}_{22})^T \end{bmatrix} \mathbf{b} = \mathbf{0}$$

If n images of the model plane are observed, by stacking n such equations shown on the previous slide, we have:

$$\mathbf{V}\mathbf{b} = \mathbf{0}$$

Where V is a $2n \times 6$ matrix. If $n \geq 3$, we will have in general a unique solution b defined up to a scale factor. The solution to this equation is well known as the eigenvector of VT V associated with the smallest eigenvalue (equivalently, the right singular vector of V associated with the smallest singular value). Once b is estimated, we can compute the intrinsic and extrinsic parameter of the camera.

6. Maximum Likelihood Estimation

The above solution is obtained through minimizing an algebraic distance, which is not physically meaningful. We can refine it through maximum likelihood inference.

We are given n images of a model plane and there are m points on the model plane. Assume that the image points are corrupted by independent and identically distributed noise. The maximum likelihood estimate can be obtained by minimizing the following functional:

$$\sum_{i=1}^{n} \sum_{j=1}^{m} \|\mathbf{m}_{ij} - \hat{\mathbf{m}}(\mathbf{A}, \mathbf{R}_i, \mathbf{t}_i, \mathtt{M}_j)\|^2$$

Minimizing this is a nonlinear minimization problem, which is solved with the Levenberg-Marquardt Algorithm. It requires an initial guess of $A, \{R_i, t_i | i = 1..n\}$ which can be obtained using the closed form solution.

7. Radial Distortion Correction and Refine Estimation

A desktop camera usually exhibits significant lens distortion, especially radial distortion. The paper proposes a way to correct radial distortion.

Let (u,v) be the pixel image coordinates, and ($\breve{u}$,$\breve{v}$) the corresponding real observed image coordinates. The ideal points are the projection of the model points according to the pinhole model. Similarly, (x, y) and ($\breve{x}$, $\breve{y}$) are the ideal (distortion-free) and real (distorted) normalized image coordinates. We have

$$\breve{x} = x + x[k_1(x^2+y^2) + k_2(x^2+y^2)^2]$$

$$\breve{y} = y + y[k_1(x^2+y^2) + k_2(x^2+y^2)^2],$$

where k_1 and k_2 are the coefficients of the radial distortion. The center of the radial distortion is the same as the principal point. From* $\breve{u} = u_0 + \alpha\breve{x} + \gamma\breve{y}$ and $\breve{v} = v_0 + \beta\breve{y}$ and assuming $\gamma = 0$, we have

$$\breve{u} = u + (u-u_0)[k_1(x^2+y^2) + k_2(x^2+y^2)^2]$$

$$\breve{v} = v + (v-v_0)[k_1(x^2+y^2) + k_2(x^2+y^2)^2]$$

Estimating Radial Distortion by Alternation. As the radial distortion is expected to be small, one would expect to estimate the other five intrinsic parameters, using the technique described in Sect. 3.2, reasonable well by simply ignoring distortion. One strategy is then to estimate k_1 and k_2 after having estimated the other parameters, which will give us the ideal pixel coordinates (u, v). Then, we have two equations for each point in each image:

$$\begin{bmatrix} (u-u_0)(x^2+y^2) & (u-u_0)(x^2+y^2)^2 \\ (v-v_0)(x^2+y^2) & (v-v_0)(x^2+y^2)^2 \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \begin{bmatrix} \breve{u}-u \\ \breve{v}-v \end{bmatrix}$$

Given m points in n images, we can stack all equations together to obtain in total 2mn equations. Dk = d, where $k = [k_1, k_2]^T$. The linear least-squares solution is given by

$$k = (D^T D)^{-1} D^T d$$

A new maximum likelihood estimate can be obtained by minimizing the following functional with radial distortion factors k1 and k2:

$$\sum_{i=1}^{n}\sum_{j=1}^{m} \|\mathbf{m}_{ij} - \breve{\mathbf{m}}(\mathbf{A}, k_1, k_2, \mathbf{R}_i, \mathbf{t}_i, \mathtt{M}_j)\|^2$$

If the model plane at the second position is parallel to its first position, then the second homography does not provide additional constraints. For known pure translation, the author gives us a solution.

8. Experiment results

For computer simulated experiment

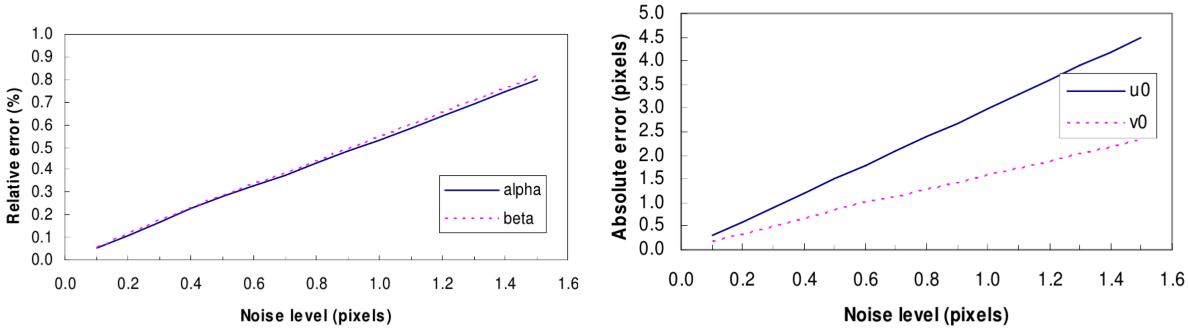


Figure 1: Errors vs. the noise level of the image points

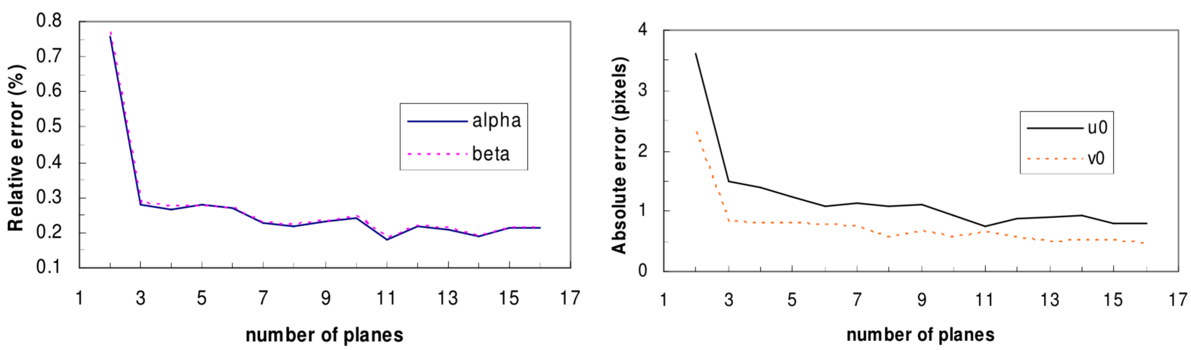


Figure 2: Errors vs. the number of images of the model plane

For real images data

Table 1: Results with real data of 2 through 5 images

nb	2 images			3 images			4 images			5 images		
	initial	final	σ	initial	final	σ	initial	final	σ	initial	final	σ
α	825.59	830.47	4.74	917.65	830.80	2.06	876.62	831.81	1.56	877.16	832.50	1.41
β	825.26	830.24	4.85	920.53	830.69	2.10	876.22	831.82	1.55	876.80	832.53	1.38
γ	0	0	0	2.2956	0.1676	0.109	0.0658	0.2867	0.095	0.1752	0.2045	0.078
u_0	295.79	307.03	1.37	277.09	305.77	1.45	301.31	304.53	0.86	301.04	303.96	0.71
v_0	217.69	206.55	0.93	223.36	206.42	1.00	220.06	206.79	0.78	220.41	206.59	0.66
k_1	0.161	−0.227	0.006	0.128	−0.229	0.006	0.145	−0.229	0.005	0.136	−0.228	0.003
k_2	−1.955	0.194	0.032	−1.986	0.196	0.034	−2.089	0.195	0.028	−2.042	0.190	0.025
RMS	0.761	0.295		0.987	0.393		0.927	0.361		0.881	0.335	

According to the experiment results, the sample deviations for all parameters are quite small, which implies that the proposed algorithm is quite stable.

9. My assessment

Critics for the paper

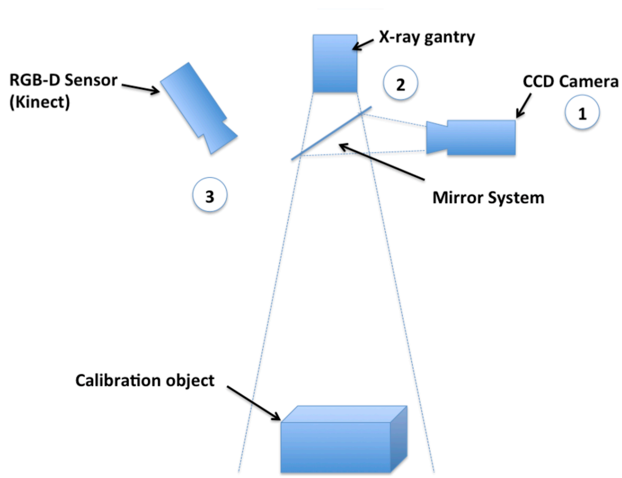
Pros

- Provide a flexible, robust and low cost calibration techniques
- Explain the math clearly
- Complete experiments

Cons

- Does not explain the maximum likelihood estimation method clearly
- Does not method about multiple cameras calibration.

Relation to our application



For our system setup, if we perform the calibration for CCD camera and Kinect RGB camera, we can get the intrinsic and extrinsic parameters of them, therefore, we can estimate a transformation from the Kinect RGB camera coordinate to CCD camera coordinate. Then we can assign depth data from the Kinect sensor to the pixels of CCD camera.