Camera Self-Calibration for the ARToolKit

J. Abdullah and K. Martinez
Intelligence, Agents and Multimedia, University of Southampton
ja100r@ecs.soton.ac.uk, km@ecs.soton.ac.uk

Abstract

Camera calibration is an essential and important part of an Augmented Reality (AR) system. The use of a plane-based calibration technique can give a good accuracy, which can be important for AR applications. The calibration technique used in the current ARToolKit requires user intervention, which is prone to error and involves a lengthy calibration time. The camera has to be recalibrated every time the focal length changes which is cumbersome and less suitable for applications where a more automated and easier approach is needed. This paper investigates the use of camera self-calibration for the ARToolKit, which has the advantage of simplicity of implementation. In order to improve its accuracy, a distortion model is also investigated. In this context several interesting results are presented.

1. Introduction

1.1 Augmented Reality

Augmented Reality (AR) is actually overlaying 3D virtual objects on the real world in real time. AR enhances the user's view of the real world with visual information from the computer. According to [8], AR is a system that has the following three characteristics:

1) Combines real and virtual
2) Interactive in real time
3) Registered in 3-D

Therefore, overlaying 2D virtual objects on the real world could not be considered AR. Films like "Jurassic Park" also could not be regarded as AR since they are not interactive media.

Camera calibration is one of the important tasks in AR systems. It establishes the projection from the 3D world co-ordinates to the 2D image co-ordinates. This can be achieved by finding the intrinsic and extrinsic camera parameters. Once the parameters are determined, 3D information can be inferred from 2D information, and vice versa.

Regardless of the specific application, existing camera calibration techniques can be classified into two [9] techniques. This classification divides camera calibration into situations with known scenes and unknown scenes. Normally, for known scenes a planar pattern is used to find the camera parameters whereas for unknown scenes, camera motion is used. For the case where neither the scene nor the camera motion is known, it requires self-calibration.

The most likely problem faced in AR applications is the registration problem [8]. The virtual object must be properly aligned with the real world. There are two types of error source that can cause registration problems: static and dynamic [10]. Static errors are the ones that cause registration errors when both the user's viewpoint and the object in the environment remains still. Dynamic errors are the ones that cause registration errors when either the viewpoint or the objects begin to move.

One way to minimize the registration errors is by having a robust camera calibration. In order to achieve this, certain criteria must be met. Tsai [1] stated that the calibration procedure should be an autonomous process, meeting certain accuracy requirements, reasonably efficient and versatile. Therefore, any work to develop more robust calibration techniques should have at least the aforementioned criteria in mind.

1.2 The ARToolKit

The ARToolKit (Augmented Reality ToolKit) is a software useful in creating AR applications. Three types of AR display that can be used with the ARToolKit including monitor based, video see-through and optical see-through displays. Different kinds of displays require different kinds of calibration routines. The calibration procedure in the ARToolKit can be categorised as semi-automatic calibration since it involves users intervention as well as the computer.

To calibrate a camera for monitor based display, the ARToolKit uses two different patterns. The first pattern is used to estimate the image centre point and lens distortion. The second pattern is used to estimate other camera parameters. The users' intervention is needed and some procedures need to be done repetitively. This can be prone to error, cumbersome and time consuming

especially when the focal length changes either intentionally or not.

In this paper, we propose a self-calibration technique for the camera in a monitor-based AR display for the ARToolKit. Our aim is to find parameters in the 3x3 calibration matrix. Once these parameters are determined, they can be used to calculate the transformation matrix, which is important to make sure the virtual object is overlaid accurately.
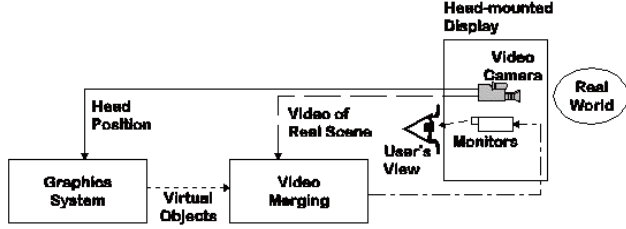

Figure 1: Monitor-based AR display

2. Background theory

2.1 Calibration matrix

We consider the pinhole model for the camera model. The camera is described by the following 3x3 upper triangular matrix:

$$\mathbf{K} = \begin{bmatrix} \alpha_u & r & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{1}$$

where α_u and α_v are the horizontal and vertical scale factors respectively, r is the degree of slant between the horizontal and vertical image axes, and (u_0, v_0) is the principal point of the image. In this paper, we use a reduced set of intrinsic parameters. We will assume that $r = 0$ which means that the image axes are orthogonal which is almost true in practice.

2.2 Camera Model

The transformation from 3D world co-ordinates to camera pixel co-ordinates can be modelled as a process of four steps [1, 2, 4]:

Step 1:
Rigid body transformation from the object world coordinate system (X_w, Y_w, Z_w) to the camera 3D coordinate system (X, Y, Z):

$$[X \quad Y \quad Z]^T = \mathbf{R}[X_w \quad Y_w \quad Z_w]^T + \mathbf{t} \tag{2}$$

Step 2:
Transformation from 3D camera coordinates (X, Y, Z) to ideal (undistorted) image coordinates (x, y) using perspective projection with pinhole camera geometry:

$$x = f\frac{X}{Z}, \qquad y = f\frac{Y}{Z}, \tag{3}$$

where f is the effective focal length

Step 3:
Lens distortion:

$$x = \hat{x} + \delta_x, \qquad y = \hat{y} + \delta_y, \tag{4}$$

where $(\hat{x}, \hat{y})$ are the distorted or true image coordinates on the image plane, and (δ_x, δ_y) are the distortion corrections to (x, y). The value for δ_x and δ_y will be discussed later.

Step 4:
Affine transformation from real image coordinates $(\hat{x}, \hat{y})$ to frame buffer (pixel) image coordinate (u, v):

$$u = d_x^{-1}\hat{x} + u_0, \qquad v = d_y^{-1}\hat{y} + v_0, \tag{5}$$

where (u_0, v_0) are the coordinates of the image centre (the principal point) in the frame buffer, d_x and d_y are the distance between adjacent pixels in the horizontal and vertical directions of the image plane, respectively.

Lens distortion mainly comprises two components: radial and decentering. Radial distortion is caused by imperfect lens shape and decentering is usually caused by improper lens assembly. The distortion corrections (δ_x and δ_y) are expressed as power series in radial distance r:

$$\delta_x = \hat{x}(k_1 r^2 + k_2 r^4 + k_3 r^6 + \cdots)$$
$$+ [p1(r^2 + 2\hat{x}^2) + 2p_2\hat{x}\hat{y}](1 + p_3 r^2 + \cdots) \tag{6}$$

$$\delta_y = \hat{y}(k_1 r^2 + k_2 r^4 + k_3 r^6 + \cdots)$$
$$+ [2p_1\hat{x}\hat{y} + p_2(r^2 + 2\hat{y}^2)](1 + p_3 r^2 + \cdots) \tag{7}$$

where $r = \sqrt{\hat{x}^2 + \hat{y}^2}$, k_1 , k_2 and k_3 are coefficients of radial distortion, and p_1 , p_2 and p_3 are coefficients of decentering distortion.

In order to keep the algorithm simpler for the ARToolKit, we only include radial distortion. According to [1], unless one is specifically concerned with the reduction of distortion to very low levels, it is likely that the distortion function is totally dominated by the radial components, and especially dominated by the k_1 term.

2.3 Epipolar Geometry

Epipolar geometry can be described by a 3x3 matrix called the fundamental matrix and it has the following expression

$$\mathbf{F} \cong \mathbf{K}^{-T}\mathbf{S(t)}\mathbf{R}\mathbf{K}^{-1} \tag{8}$$

where $\mathbf{R}$ and $\mathbf{t}$ represent the relative rotation and translation between the two camera positions, $\mathbf{S(t)}$ is the skew-symmetric matrix associated with three-vector $\mathbf{t}$. $\cong$ stands for equality up to a scale factor.

2.4 Fundamental Matrix Estimation

The fundamental matrix can be robustly estimated from point matches given at least seven point-to-point correspondence. If a point $\mathbf{m}_i = \begin{bmatrix} u_i & v_i \end{bmatrix}^T$ in the first image is matched to a point $\mathbf{m}'_i = \begin{bmatrix} u'_i & v'_i \end{bmatrix}^T$ in the second image, they must satisfy the following epipolar equation:

$$\widetilde{\mathbf{m}}'^T_i \mathbf{F} \widetilde{\mathbf{m}}_i = 0 \tag{9}$$

In this paper we use publicly available software "image matching" [12] to get the accurate determination of point correspondence and to estimate the fundamental matrix.

3. Self Calibration

There are several techniques reported in the literature for self-calibration [6, 7]. These techniques are all rely on known motions of the cameras. In [7] the motion of the camera is assumed to be translational, while in [6] rotational motions of the camera are considered, but rotation must be through known angles.

The method chosen here is the algebraic Dornaika's method [5]. This method is used because it outperforms the Kruppa equations in terms of convergence and accuracy and also for its simpler algorithm, which is more suitable for the ARToolKit.

The method is derived from equation (8). The equation can be rewritten as

$$\mathbf{K}^T\mathbf{FKR}^T \cong \mathbf{S(t)} . \tag{10}$$

Let $a_{lj}, 1 \le l, j \le 3$ be entries of $\mathbf{K}^T\mathbf{FKR}^T$. From this it follows that

$$a_{11} = a_{22} = a_{33} = a_{12} + a_{21} = a_{13} + a_{31} = a_{23} + a_{32} = 0 \tag{11}$$

The rotation $\mathbf{R}$ is represented by its associated unit quaternion $\mathbf{q} = (q_0, q_x, q_y, q_z)^T$. More details of the derivation of the method can be found in [5].

For n fundamental matrix, we accumulate the constraints (11) by building a positive error function f that will be minimized over the unknown. The error function can have the following form:

$$f(\alpha_u, \alpha_v, u_o, v_o, \mathbf{q}_1, \ldots, \mathbf{q}_n) = \sum_{i=1}^{n} \left\| \mathbf{v}_i \right\|^2 + \lambda(1 - \left\| \mathbf{q}_i \right\|^2)^2] \tag{12}$$

where $\mathbf{v} = (a_{11}, a_{22}, a_{33}, a_{12} + a_{21}, a_{13} + a_{31}, a_{23} + a_{32})^T$, and λ is a real positive number.

3.1 Distortion

There are two ways of dealing with distortion. As a start we are concerned with the radial distortion. Often in literature, the fundamental matrix is estimated based on the assumption that distortion in the captured image is corrected offline. However, there is a paper reported by Zhang [4], which integrates the distortion correction when estimating the epipolar geometry. In this context we only concentrate on the former method for its simplicity.

4. Experiment

An experiment was conducted in order to check the accuracy of the intrinsic parameters from the camera self-calibration routine. The results are compared with the one from the ARToolKit. In the experiments carried out to test the self-calibration algorithm, image sequences, rather than individual image were used.

$\mathbf{K}$ is initialized using reasonable figures quoted by camera the manufacturer or from educated guess. The quaternion $\mathbf{q}$ is initialized through the following steps.

1. The camera that needs to be calibrated should capture at least 3 images by moving the camera.
2. Find the point correspondence between each pair of images based on relaxation and correlation techniques.
3. From the point correspondence, the fundamental matrix is estimated. Each pair will produce one fundamental matrix.
4. The essential matrix would then be estimated based on this relation $\mathbf{E}_i \cong \mathbf{K}^T\mathbf{F}_i\mathbf{K}$.
5. Based on the essential matrix, each quaternion $\mathbf{q}_i$ is estimated by factorizing the corresponding essential matrix as $\mathbf{E}_i \cong \mathbf{S}(\mathbf{t}_i)\mathbf{R}(\mathbf{q}_i)$.

The initial intrinsic parameters together with the obtained quaternions constitute the starting point of the nonlinear equation in (12). The calibration matrix from self-calibration and ARToolKit is presented as follows:
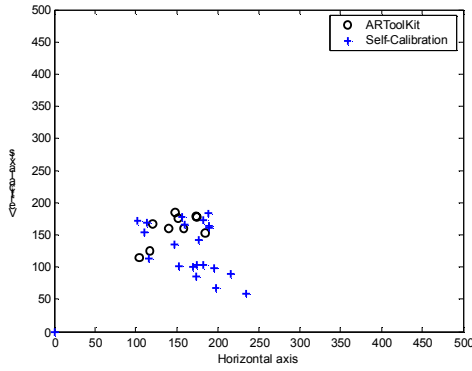
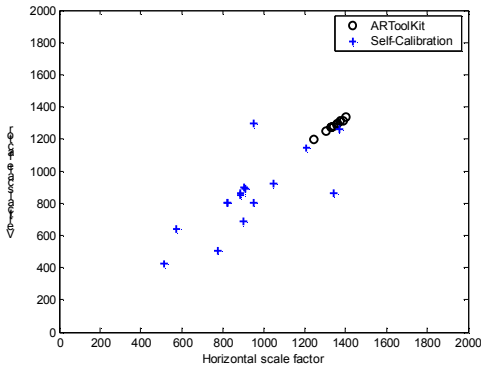**Figure 2: Principal Point without distortion correction.**

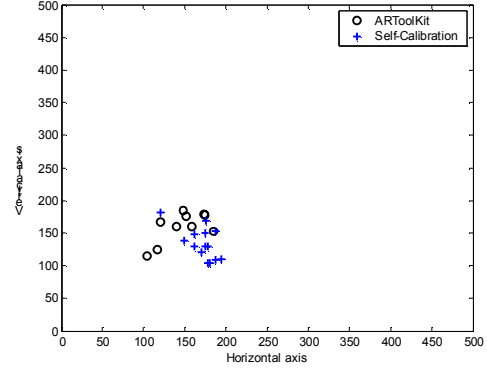**Figure 3: Scale Factor without distortion correction.**

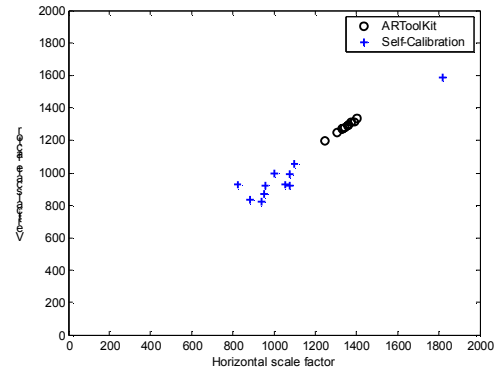**Figure 4: Principal point after distortion correction**

Figure 5: Scale factor after distortion correction

From these results, we can observe that the principal point estimation for self-calibration without distortion correction is quite reliable. However **Figure 4** and **Figure 5** show improvements in the clusters for the self-calibration method in comparison to clusters in **Figure 2** and **Figure 3**. From here we can presume that by taking distortion into account will improve the accuracy of the camera parameters.

5. Conclusion and Future Work

Although the algorithm proposed here is not as robust as the existing algorithm in the ARToolKit, but this could be considered as a first step towards a more automatic calibration process without the needs for a particular pattern. Improvement can be made by accurately estimating the fundamental matrix which is largely related to the accuracy of point correspondence matching. Once this is achieved, it is easier to obtain a good estimation of the camera parameters.

For future work, investigation on Zhang's distortion method [4] can be employed into distortion parameters calculation. This makes the calibration more automatic rather than having to correct the distortion offline.

6. References

[1] R. 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*, August 1987, pp.323-344.

[2] D.C. Brown, "Close-range camera calibration", *Photogrammetric Engineering*, 1971, pp. 855-866.

[3] C. C. Slama, "Manual of Photogrammetry", *American Society of Photogrammetry*, 1980.

[4] Zhengyou Zhang, "On the epipolar geometry between two images with lens distortion", *In Proceeding of the International. Conference on Pattern Recognition,* Vienna, August 1996, pp. 407-411.

[5] F. Dornaika and R. Chung, "An Algebraic Approach to Camera Self-Calibration", *Computer Vision and Image Understanding*, 2001, pp.195-215

[6] Anup Basu, "Active calibration: Alternative strategy and analysis", *In Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition*, 1993, pp. 495-500.

[7] Lisa Dron, "Dynamic camera self-calibration from controlled motion sequences", *In Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition*, 1993, pp. 501-506,.

[8] R.T. Azuma, "A survey of augmented reality", *Presence: Teleoperators and virtual environments*, 1997, vol. 6, no. 4, pp. 355-385,.

[9] Milan Sonka, Vaclav Hlavac, Roger Boyle*, Image Processing, Analysis, and Machine Vision*, 2nd Edition, PWS Publishing, 1999.

[10] Holloway, Richard, "Registration Errors in Augmented Reality", *Ph.D dissertation*. UNC Chapel Hill Department of Computer Science technical report TR95-016, August 1995.

[11] Hartley, R, "Projective reconstruction and invariants from multiple images", *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1994, pp. 1036-1040.

[12]http://www.sop.inria.fr/robotvis/personnel/zzhang/softwares.html