# EE 243 Advanced Computer Vision: Homework 1 Report

Tue-Cuong Dong-Si
tdongsi@ee.ucr.edu

## 1. Camera calibration

In this section, the method to calibrate a camera is presented. After the camera calibration procedure, the intrinsic parameters, namely the aspect ratio $\alpha$ and the effective focal length in pixels $f_x = f/s_x$, and the extrinsic parameters, consisting of the rotation matrix $\mathbf{R}$ and translation vector $\mathbf{t}$ will be estimated. The steps of the calibration process is described in the following subsections.

### 1.1. Measure 3D world coordinates

In camera calibration, we need to form the correspondences between the points in an image to the points in a designated world reference frame. The world reference frame is chosen such that it is easy and accessible to measure other points in that reference frame. For our particular scene, a point at the base of the tree pot is chosen as the origin of our world reference frame as shown in Fig. 1 (the green point). In addition, the z axis of this reference frame is normal to the ground plane pointing up, the y axis is along the base of the tree pot toward the camera, and the Ox is normal to both. Using a measuring tape, we measured and computed the 3D coordinates of several points in that world reference frame. We collect measurements for 28 points, and use 20 of them for calibration while the rest are used for testing and verification.

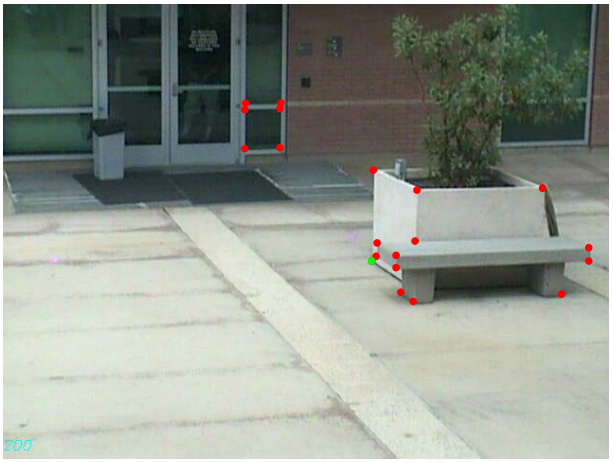

Figure 1. Selected points for calibration (red) and the designated world frame origin point {O} (green).

### 1.2. Measure image coordinates

Since the camera view is static, after having the 3D coordinates, we need to find the corresponding image coordinates (pixel) in any camera image. To find the pixel coordinates accurately, we crop a small local region around each selected point and use Harris corner detector to find the most prominent corner point in the cropped region. In addition, in our method, it is assumed that the origin of the image reference frame is at the image center. Therefore, given that the image size is $640 \times 480$, the pixel coordinates are subtracted by (320, 240), assuming the principle point is exactly at the center of the image. The radial distortion is also neglected in this calibration process.

### 1.3. Construct and solve the first linear system

Having 20 points with coordinates in both image and world reference frame, we proceed to construct the $20 \times 8$ matrix $\mathbf{A}$ described by Eq. (6.8) in [3]. From [3], we have that:

$$\mathbf{A}\mathbf{v} = \mathbf{0} \tag{1}$$

where $\mathbf{v} = [r_{21} \;\; r_{22} \;\; r_{23} \;\; t_y \;\; \alpha r_{11} \;\; \alpha r_{12} \;\; \alpha r_{13} \;\; \alpha t_x]^T$ is the vector of camera parameters that we want to estimate. In theory, $\mathbf{v}$ can be determined by finding the null vector of $\mathbf{A}$, scaled by some unknown scale factor $\gamma$. However, in practice, due to noises and measurement errors, $\mathbf{A}$ becomes full-ranked. Therefore, we have to find $\mathbf{v}$ from SVD of $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T$ and $\mathbf{v}$ is the column of $\mathbf{V}$ corresponding to the smallest singular value.

Having $\mathbf{v}$ and using the constraints for entries of rotation matrix that $r_{21}^2 + r_{22}^2 + r_{23}^2 = 1$ and $r_{11}^2 + r_{12}^2 + r_{13}^2 = 1$, we can determine $|\gamma|$ and $\alpha$, and subsequently, other parameters in $\mathbf{v}$. The sign of $\gamma$ can be determine from the test of any point that is far enough from the image center:

$$x(r_{11}X^w + r_{12}Y^W + r_{13}Z^w + T_x) > 0 \tag{2}$$

If the test returns true, then $\gamma$ is negative and we have to reverse the sign of $\mathbf{v}$. After this step, we have determined the first two rows of rotation matrix $\mathbf{R}$ and the first two components of translation vector $\mathbf{t}$.

The third row of the rotation matrix $\mathbf{R}$ is computed as the cross product of the first two rows. Due to noises and measurement error, the estimated $\hat{\mathbf{R}}$ may not be orthogonal. Therefore, to enforce orthogonality, we use SVD of $\hat{\mathbf{R}} = \mathbf{U}_r\mathbf{D}_r\mathbf{V}_r^T$ and the new $\mathbf{R}' = \mathbf{U}_r\mathbf{V}_r^T$ is the "closest" orthogonal matrix.

### 1.4. Construct and solve the second linear system

Constructing the matrix $\mathbf{A}$ and vector $\mathbf{b}$ described by Eq. (6.14) in [3], we can estimate the last component of

the translation vector $\mathbf{t}$ and the effective focal length in pixels $f_x$. At the end of the calibration process, we have estimated all extrinsic parameters and two intrinsic parameters of the camera.

### 1.5. Results

For our experiment, the calibration returns the aspect ratio as $\alpha = 1.0165$, the effective focal length $f_x = 1457$ pixels, the translation $\mathbf{t} = [-0.6774 \ -0.2572 \ 13.5129]^T$ (meters) and the rotation matrix

$$\mathbf{R} = \begin{bmatrix} 0.9684 & -0.2491 & -0.0156 \\ -0.0399 & -0.2161 & 0.9755 \\ -0.2464 & -0.9441 & -0.2192 \end{bmatrix} \quad (3)$$

To verify the calibration results, we reproject other measured 3D world points as well as the bases of the world reference frame on the image, as shown in Fig. 2 (blue points). The RMS reprojection error is $(2.7, 2.5)$ pixels.

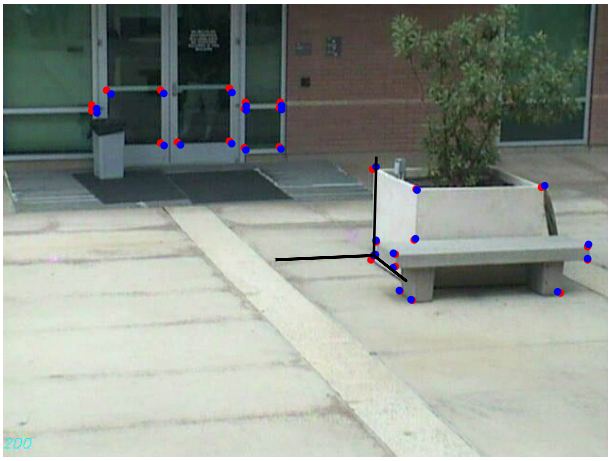

Figure 2. Calibration and re-projection results. The black lines are the orthogonal bases of the world reference frame.

## 2. 3D scene estimation

For 3d model estimation, one would use information from different camera views to reconstruct 3D world on from 2D views. Stereo methods were attempted but they failed catastrophically. Among several factors that contribute to the challenges faced by 3D model estimation of this particular scene, the major factor is that there is little overlapping between views of camera 22 and other cameras in this scene, as shown in Fig. 3. In particular, the bases and origin of the designated world reference frame in Figure. 2 are not seen in any other camera views. Different attempts to establish the approximate relative translation and attitude between camera views give unsatisfactory results. With only one available camera view, it's possible [1, 2] but very challenging to reconstruct 3D model of the scene as it requires significant prior information and training.

(a) Camera 16 (b) Camera 17

(c) Camera 27 (d) Camera 31

(e) Camera 36 (f) Camera 37

Figure 3. Different camera views of the scene 28.

## References


[1] D. Hoiem, A. Efros, and M. Hebert. Geometric context from a single image. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 1, pages 654–661. IEEE, 2005.

[2] A. Saxena, M. Sun, and A. Ng. Make3d: learning 3d scene structure from a single still image. *IEEE transactions on pattern analysis and machine intelligence*, pages 824–840, 2008.

[3] E. Trucco and A. Verri. *Introductory techniques for 3-D computer vision*, volume 93. Prentice Hall New Jersey, 1998.