CSCM77
Camera Model and Calibration

Prof. Xianghua Xie

x.xie@swansea.ac.uk

http://csvision.swan.ac.uk

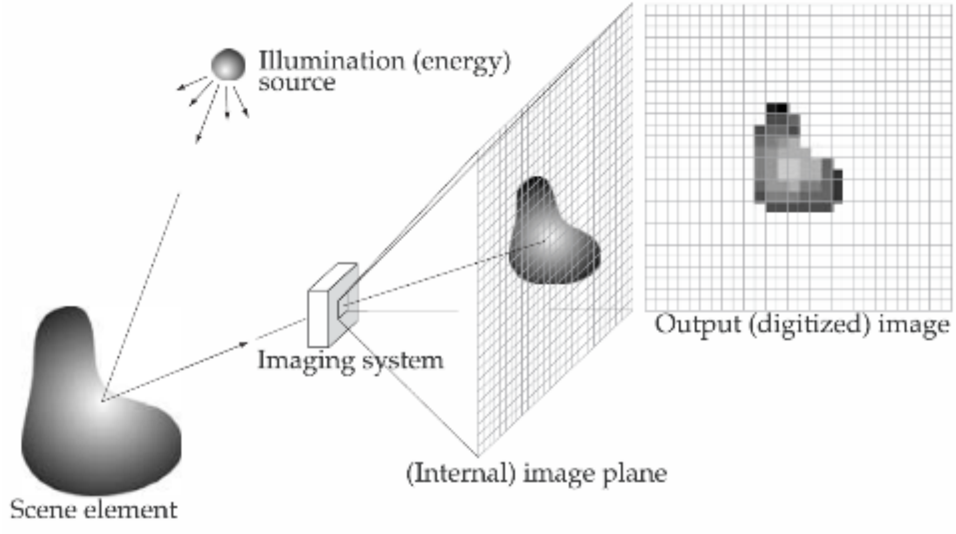
Content

- Camera models
  - Perspective or pin-hole
  - Weak perspective & orthographic projection

- Parameters
  - Extrinsic (camera position and orientation)
  - Intrinsic (relationship between pixel and image plane coordinates)

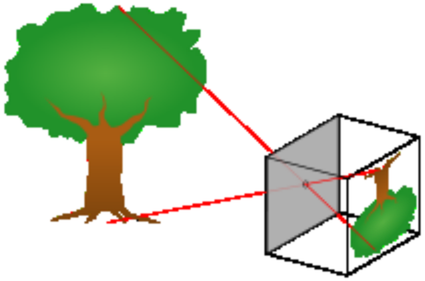
- Camera calibration
  - Calibration pattern and system
  - Direct calibration method

Digital image formation



- Light focused onto image plane using optical system
- Intensity variation across image plane captured by array of photosensitive devices, e.g. CCD
- Intensity values captured at given instant in time to give array of pixel values
- Each pixel has a position in **pixel coordinates**, e.g. (i,j) that is (i-th row and j-th column) and an intensity value E(i,j)
- Intensity values usually in the range [0,255] for monochrome
- Spatial resolution determined by effective area of image plane occupied by each pixel, s_x X s_y

Perspective pin-hole camera model



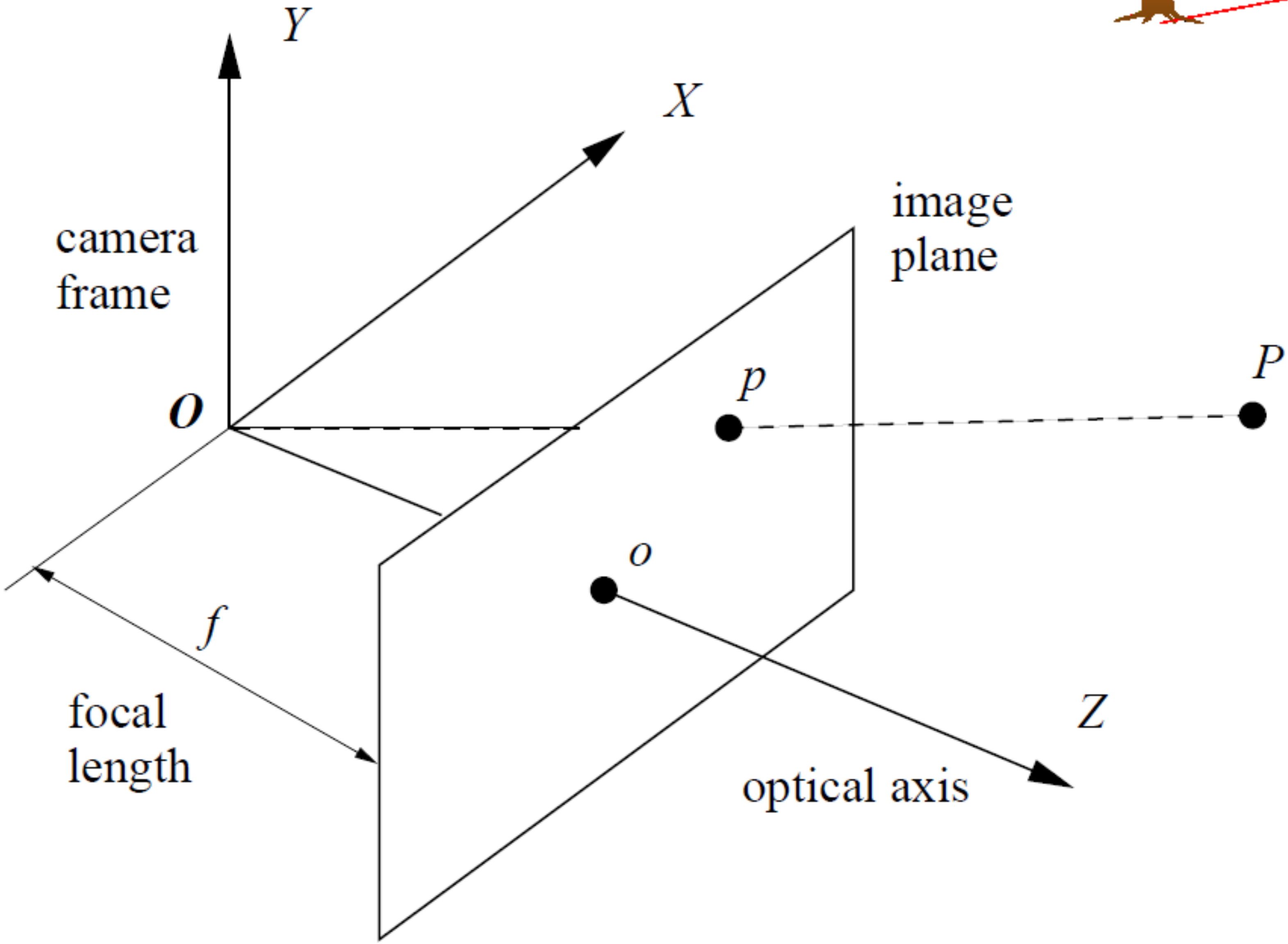
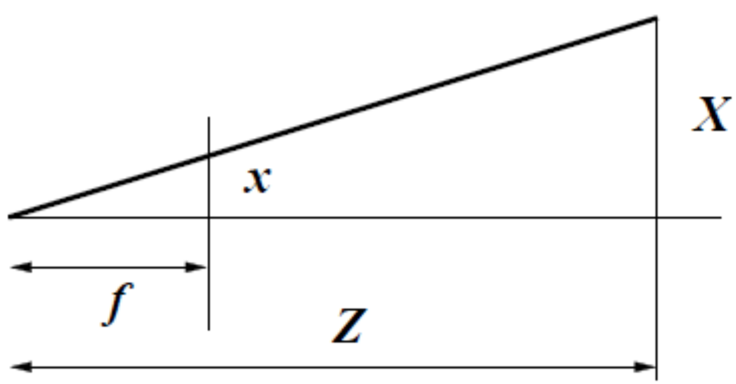


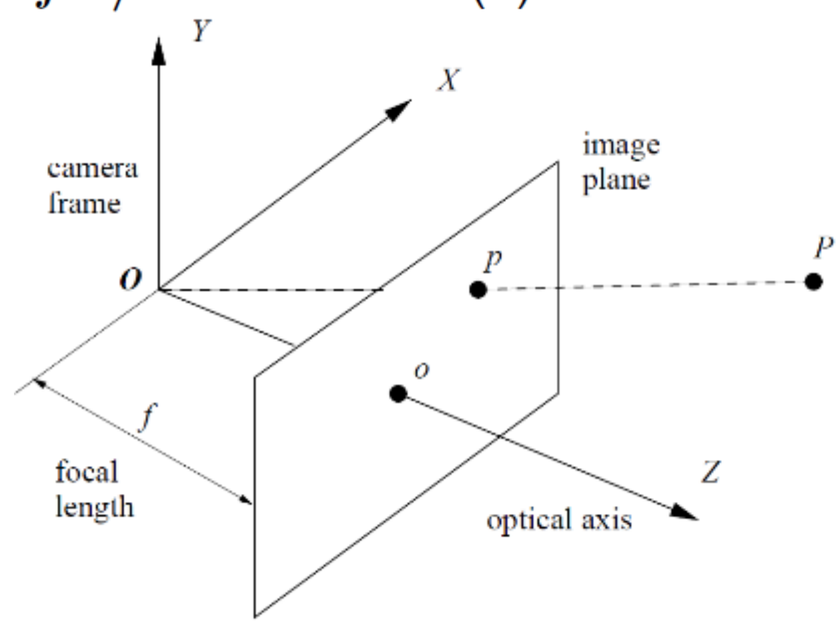
Image credit: Wiki, Calway & Fua

Perspective projection

- Assume all points defined with respect to camera frame
- Point $\mathrm{P} = [\mathrm{X,Y,Z}]^T$ in 3D space
- P projects to point $p = [x,\ y, f]^T$ in image plane
  - f is **focal length:** the parameter of interest in perspective proj.
  - T denotes transpose
- Relationship between P and p is defined as:

$$x = fX/Z \qquad y = fY/Z \qquad (1)$$





Weak perspective

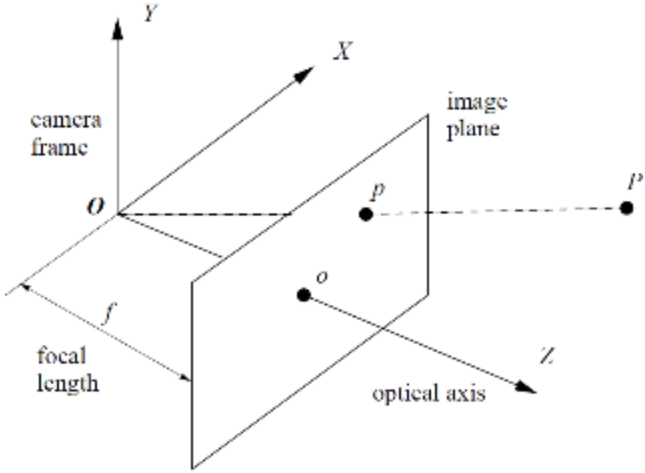
- Let maximum difference in depth (distance to camera centre, focal point) between any two scene points = δZ
- Let average depth of all scene point = $\bar{Z}$
- If the variation in depth is negligible compared to average depth, then for each scene point:

$$x \approx fX/\bar{Z} \qquad\qquad y \approx fY/\bar{Z}$$

- Since focal length and average depth are fixed constants, it becomes a linear camera model
- Viable approximation when variation in depth is less than 5%
- Known as "weak perspective": retains some sense of perspective
- If ignore the constant values, it becomes the orthographic projection (x=X, y=Y): no perspective

Camera parameters

- Previously, points are defined with respect to camera reference frame, e.g. perspective camera model
- However, we require
  - Camera reference frame can be located with respect to some **world reference**
  - Establish the relationship between the coordinates of the **image plane** and **pixel coordinates**

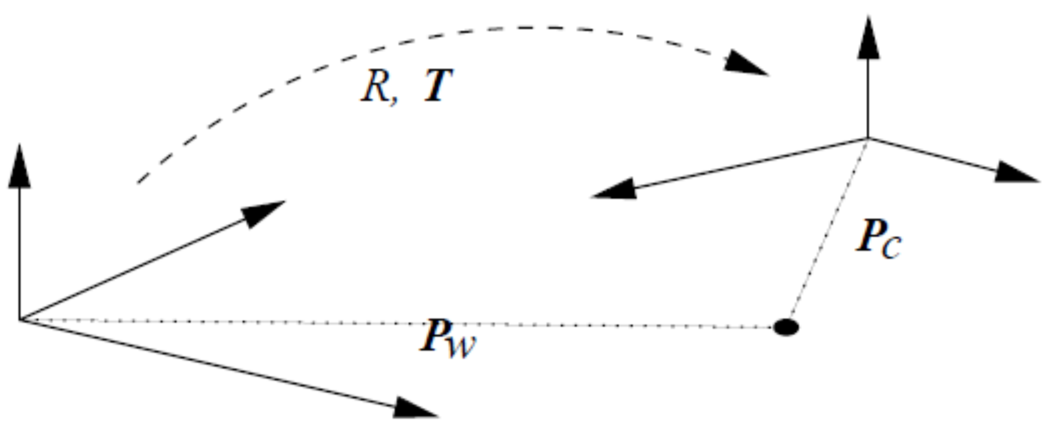


  - Working out the above two relationships require two sets of camera parameters
    - Extrinsic parameters
    - Intrinsic parameters

Extrinsic camera parameters

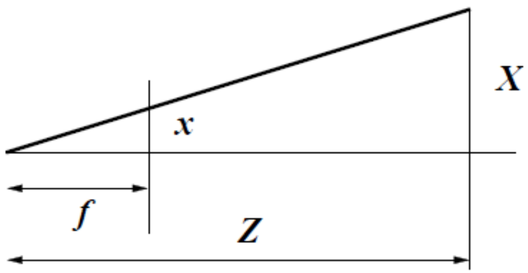
- Parameters that define location and orientation of camera reference frame with respect to known world reference frame
- Camera and world frame coordinates related by
  - 3 by 3 rotation matrix R, which defines rotations about X, Y and Z
  - 3D translation vector T, which defines translation between origins

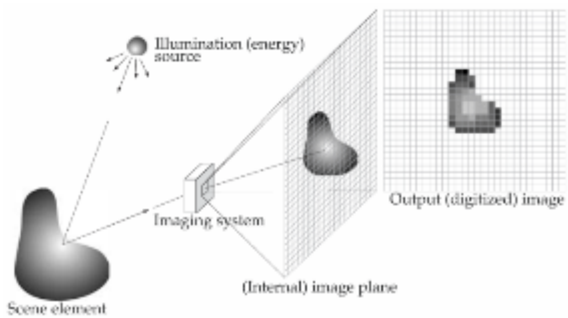
$$\mathrm{P}_c = \mathrm{R}\mathrm{P}_w + \mathrm{T} \qquad (2)$$



Intrinsic camera parameters

- For pin hole camera, 3 sets of intrinsic parameters specifying
  - The perspective projection: focal length f
  - Transformation between image plane and pixel coordinates
  - Geometric distortion introduced by optics

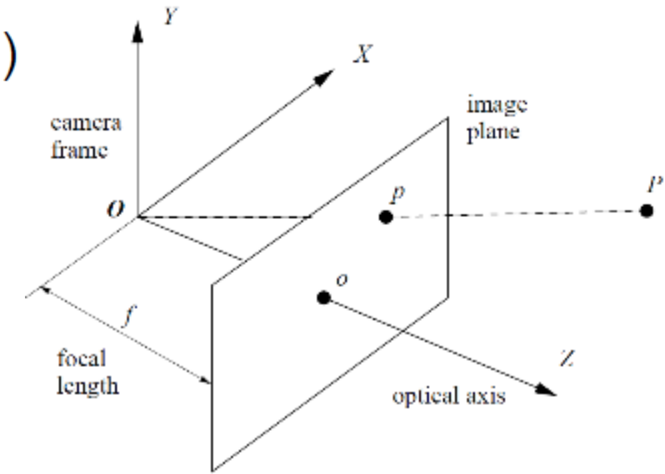




- Image plane (x,y) and pixel coordinates (x_{im}, y_{im}):

$$x = (x_{im} - o_x)s_x \qquad y = (y_{im} - o_y)s_y \quad (3)$$

  - o_x, o_y origin of image plane
  - s_x, s_y spatial resolution



Intrinsic camera parameters

- Geometric distortion can be modelled as displacements increasing with radial distance form image centre:

$$x = x_d(1 + k_1 r^2 + k_2 r^4) \qquad y = y_d(1 + k_1 r^2 + k_2 r^4)$$

  - (x_d, y_d): coordinates of distorted points
  - $r^2 = x_d^2 + y_d^2$

Linking pixel and world coordinates

- Given intrinsic and extrinsic parameters, we can link pixel and world coordinates together, i.e. mapping between pixel and world reference coordinates

- Intrinsic parameters allow us to map pixel coordinates to image plane/camera coordinates
- Perspective projection equation tells us how 3D points projected to image plane; those 3D points are defined with respect to camera coordinates
- Extrinsic parameters allow us to map the world reference and camera coordinates (rotation and translation)

Linking pixel and world coordinates

- Link pixel and world coordinates together
- Recall (1-3):
  - (1): perspective projection
  - (2): rotation and translation between two coordinates (ext.)
  - (3): image plane and pixel coordinates (int.)

$$x = fX/Z \qquad\qquad y = fY/Z \qquad\qquad (1)$$

$$\mathrm{P}_c = \mathrm{R}\mathrm{P}_w + \mathrm{T} \qquad\qquad (2)$$

$$x = (x_{im} - o_x)s_x \qquad y = (y_{im} - o_y)s_y \qquad (3)$$

- We then have:

$$(x_{im} - o_x)s_x = f\frac{\mathrm{R}_1^T \mathrm{P}_w + T_x}{\mathrm{R}_3^T \mathrm{P}_w + T_z}$$
$$(y_{im} - o_y)s_y = f\frac{\mathrm{R}_2^T \mathrm{P}_w + T_y}{\mathrm{R}_3^T \mathrm{P}_w + T_z} \qquad (4)$$

  - R_i^T is i-th row of matrix R and $\mathrm{T}=[T_x, T_y, T_z]^T$

Projection matrix

- Equation (4) can be re-arranged:

$$\mathrm{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \mathrm{M}_{int}\mathrm{M}_{ext} \begin{bmatrix} \mathrm{P}_w \\ 1 \end{bmatrix}$$

  – Where $x_{im} = x_1/x_3,\, y_{im} = x_2/x_3$

  – $\mathrm{M}_{int} = \begin{bmatrix} f/s_x & 0 & o_x \\ 0 & f/s_y & o_y \\ 0 & 0 & 1 \end{bmatrix}$

  – $\mathrm{M}_{ext} = [\mathrm{R}, \mathrm{T}] \quad (3 \times 4)$

- The 3 by 4 matrix $M = M_{int}M_{ext}$ is known as the projection matrix

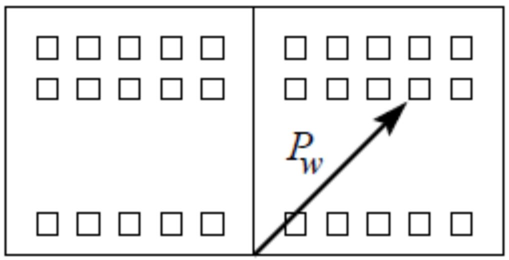
Camera calibration

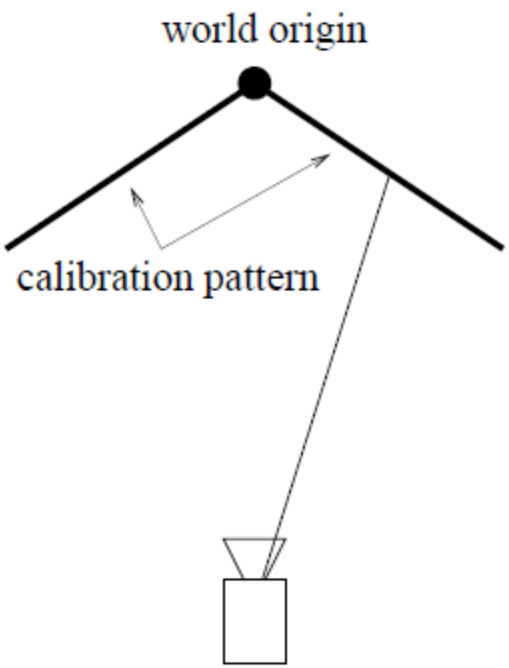
- Of course, camera intrinsic and extrinsic parameters are not known, i.e. the projection matrix is actually not determined
- However, we can use this projection matrix to estimate intrinsic and extrinsic parameters
- This is known as camera calibration

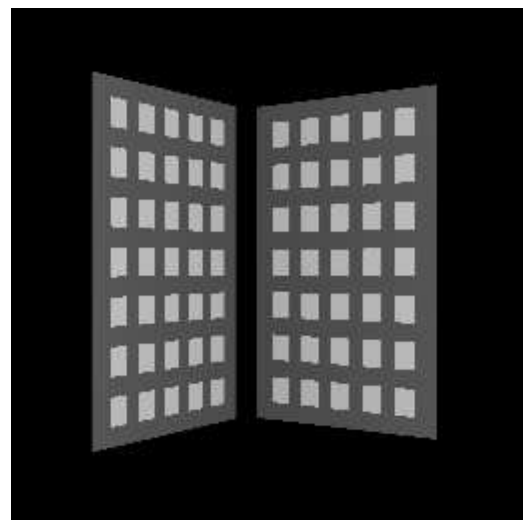
- Direct parameter calibration
  - Capture images containing calibration patterns with known 3D points
  - Determine pixel coordinates of projected points
  - Solve equations linking world and pixel coordinates using N such points and Least-Square optimisation

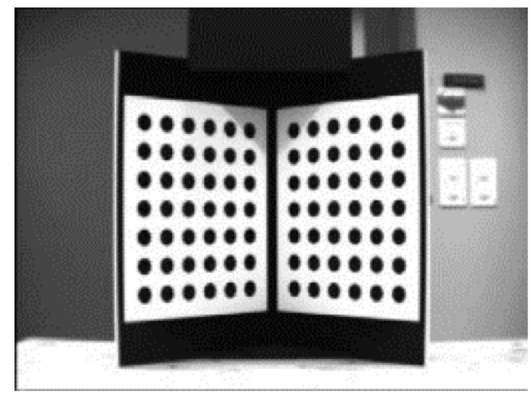
Camera calibration

- Calibration pattern is stationary
- Its origin is considered as the world reference origin









Camera calibration

- Assume image centre (o_x, o_y) is known, so

$$x'_{im} = x_{im} - o_x \qquad y'_{im} = y_{im} - o_y$$

- According to (4) we have

$$x'_{im} = f_x \frac{\mathrm{R}_1^T \mathrm{P}_w + T_x}{\mathrm{R}_3^T \mathrm{P}_w + T_z} \qquad y'_{im} = f_y \frac{\mathrm{R}_2^T \mathrm{P}_w + T_y}{\mathrm{R}_3^T \mathrm{P}_w + T_z} \qquad (5)$$

  – Where $f_x = f / s_x$ and $f_y = f / s_y$

- Hence from common denominator

$$x'_{im}(\mathrm{R}_2^T \mathrm{P}_w + T_y) - y'_{im}\alpha(\mathrm{R}_1^T \mathrm{P}_w + T_x) = 0$$

  – Where aspect ratio $\alpha = f_x / f_y = s_y / s_x$

- Reminder:

$$(x_{im} - o_x)s_x = f \frac{\mathrm{R}_1^T \mathrm{P}_w + T_x}{\mathrm{R}_3^T \mathrm{P}_w + T_z}$$

$$(y_{im} - o_y)s_y = f \frac{\mathrm{R}_2^T \mathrm{P}_w + T_y}{\mathrm{R}_3^T \mathrm{P}_w + T_z} \qquad (4)$$

Camera calibration

$$x'_{im}(\mathrm{R}_2^T \mathrm{P}_w + T_y) - y'_{im}\alpha(\mathrm{R}_1^T \mathrm{P}_w + T_x) = 0$$

- Thus, for N points above formulation will give us N linear equations with 8 unknowns:

$$[\mathrm{R}_2^T, T_y, \alpha \mathrm{R}_1^T, \alpha T_x]$$

- Solving this linear system (for example, using Singular Value Decomposition) and using constraints such as the length of rotation vectors should be one will give us

$$\mathrm{R}_1, \mathrm{R}_2, T_x, T_y, \alpha$$

Camera calibration

- Rotation matrix has to be orthogonal: $\mathrm{R}_3 = \mathrm{R}_1 \times \mathrm{R}_2$ (cross product)

- Finally, Tz and fx (and fy from aspect ratio) can be similarly worked out using equation (5) again, for example use least-square optimisation

$$x'_{im}\mathrm{R}_3^T\mathrm{P}_w + x'_{im}T_z = f_x(\mathrm{R}_1^T\mathrm{P}_w + T_x)$$

- Reminder:

$$x'_{im} = f_x\frac{\mathrm{R}_1^T\mathrm{P}_w + T_x}{\mathrm{R}_3^T\mathrm{P}_w + T_z} \qquad y'_{im} = f_y\frac{\mathrm{R}_2^T\mathrm{P}_w + T_y}{\mathrm{R}_3^T\mathrm{P}_w + T_z} \qquad (5)$$

Application

- Work such as this one extensively involves camera parameter estimation (in a very large scale)