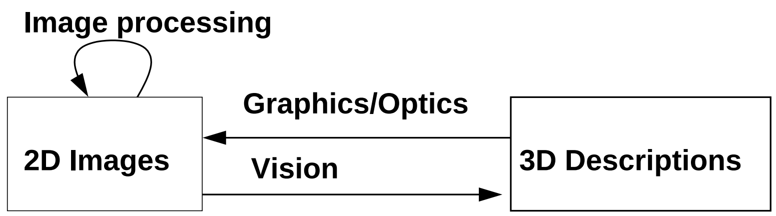
Projective geometry, camera models and calibration

Subhashis Banerjee
Dept. Computer Science and Engineering
IIT Delhi
email: suban@cse.iitd.ac.in

January 16, 2008

1 The main problems in computer vision



Correspondence problem: Match image projections of a 3D configuration.

Reconstruction problem: Recover the structure of the 3D configuration from image projections.

Re-projection problem: Is a novel view of a 3D configuration consistent with other views? (Novel view generation)

All of these require camera calibration in some form.

2 An infinitely strange perspective

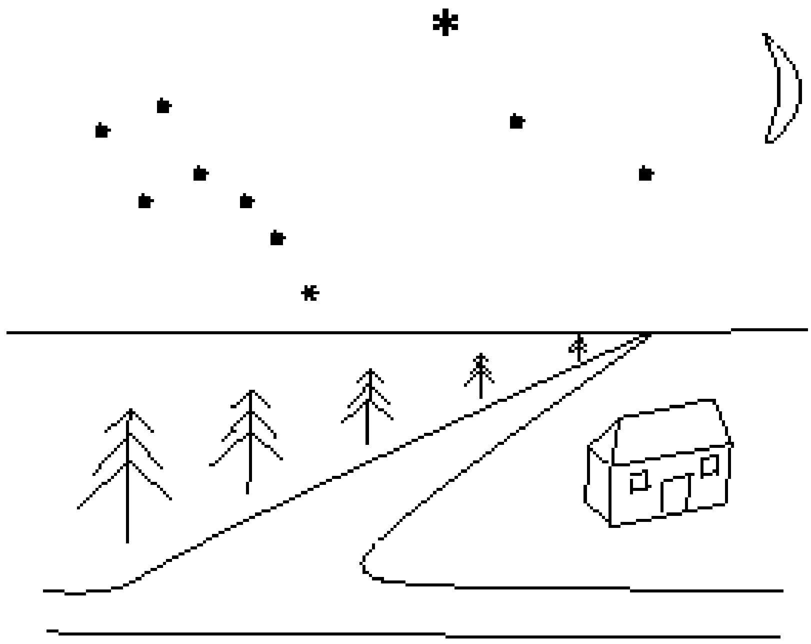


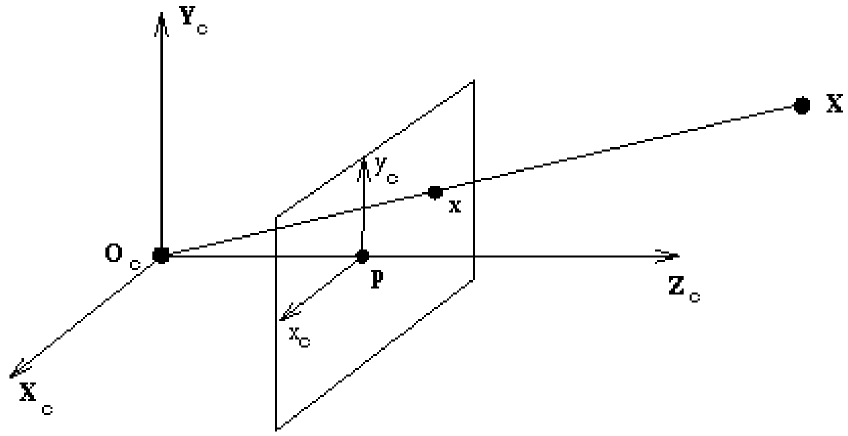
Figure 1.1: Landscape with horizon

- Parallel lines in 3D space converge in images.
- The line of the horizon is formed by 'infinitely' distant points (vanishing points).

- Any pair of parallel lines meet at a point on the horizon corresponding to their common direction.
- All 'intersections at infinity' stay constant as the observer moves.

The effects can be modelled mathematically using the 'linear perspective' or a 'pin-hole camera' (realized first by Leonardo?)

3 The pin-hole camera model



3.1 Standard perspective projection

If the world coordinates of a point are (X, Y, Z) and the image coordinates are (x, y), then

$$x = fX/Z \text{ and } y = fY/Z$$

The model is non-linear.

3.2 In terms of projective coordinates

$$\lambda \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \in \mathcal{P}^2 \text{ and } \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \in \mathcal{P}^3$$

are **homogeneous coordinates**.
The model is linear in projective geometry.

4 Basics of Projective Geometry

4.1 Affine and Euclidean geometries

- Given a coordinate system, n-dimensional real **affine space** is the set of all points parameterized by $\mathbf{x} = (x_1, \ldots, x_n)^t \in \mathbb{R}^n$.
- An affine transformation is expressed as

$$\mathbf{x}' = \mathbf{A}\mathbf{x} + \mathbf{b}$$

where $\mathbf{A}$ is a $n \times n$ (usually) non-singular matrix and $\mathbf{b}$ is a $n \times 1$ vector representing a translation.

- In the special case of when $\mathbf{A}$ is a rotation (i.e., $\mathbf{A}\mathbf{A}^t = \mathbf{A}^t\mathbf{A} = \mathbf{I}$, the the transformation is *Euclidean*.
- Transformation of one point (or one axis) completely determines an Euclidean transformation, an affine transformation in n dimensions is completely determined by a mapping of $n + 1$ points (3 points for a plane).
- It is easy to verify that an affine transformation preserves parallelism and ratios of lengths along parallel directions. In fact, coordinates in an affine geometry are defined in terms of these fundamental invariants. An Euclidean transformation, in addition to the above, also preserves lengths and angles.
- *Since an affine (or Euclidean) transformation preserves parallelism it cannot be used to describe a pinhole projection.* We need to *projective geometry* to represent such transformations.

4.2 Spherical geometry

Before we introduce *projective geometry* let us briefly consider *spherical geometry* ($\mathcal{S}^2$), which is the geometry on the surface of a sphere.

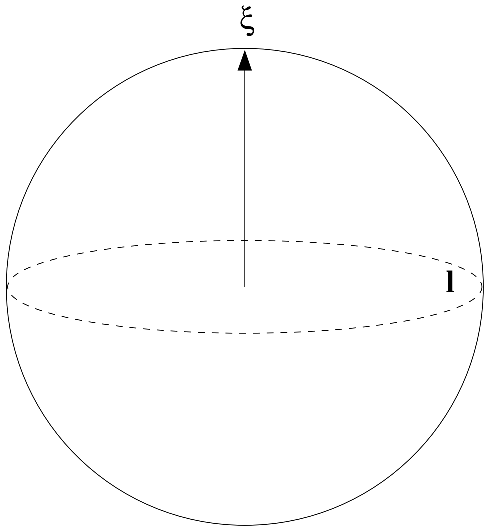
- **The space $\mathcal{S}^2$:**

$$\mathcal{S}^2 = \left\{\mathbf{x} \in \mathbb{R}^3 : ||x|| = 1\right\}$$

- **lines in $\mathcal{S}^2$:** If one begins at a point in $\mathcal{S}^2$ and travels straight ahead on the surface, one will trace out a *great circle*. Viewed as a set in $\mathbb{R}^3$ this is the intersection of $\mathcal{S}^2$ with a plane through the origin. We will call this great circle a line in $\mathcal{S}^2$:
Let ξ be a unit vector. Then,

$$\mathbf{l} = \left\{\mathbf{x} \in \mathcal{S}^2 : \xi^t\mathbf{x} = 0\right\}$$

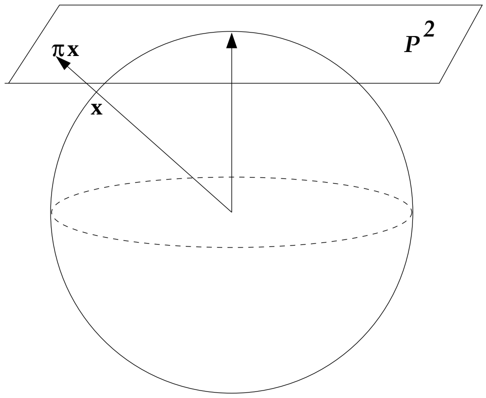
is the line with pole ξ.



- Two points $\mathbf{p}$ and $\mathbf{q}$ are *antipodal* if $\mathbf{p} = -\mathbf{q}$.
- Lines in $\mathcal{S}^2$ cannot be parallel. Any two lines intersect at a pair of antipodal points.
- A point on a line:
$$\mathbf{l} \cdot \mathbf{x} = 0 \text{ or } \mathbf{l}^T\mathbf{x} = 0 \text{ or } \mathbf{x}^T\mathbf{l} = 0$$
- Two points define a line:
$$\mathbf{l} = \mathbf{p} \times \mathbf{q}$$
- Two lines define a point:
$$\mathbf{x} = \mathbf{l} \times \mathbf{m}$$

4.3 Projective geometry

- *The projective plane $\mathcal{P}^2$ is the set of all pairs $\{\mathbf{x}, -\mathbf{x}\}$ of antipodal points in $\mathcal{S}^2$.*
- Two alternative definitions of $\mathcal{P}^2$, equivalent to the preceding one are
  1. The set of all lines through the origin in $\mathbb{R}^3$.
  2. The set of all equivalence classes of ordered triples (x_1, x_2, x_3) of numbers (i.e., vectors in $\mathbb{R}^3$) not all zero, where two vectors are equivalent if they are proportional.
- The space $\mathcal{P}^2$ can be thought of as the infinite plane tangent to the space $\mathcal{S}^2$ and passing through the point $(0, 0, 1)^t$.



- Let $\pi : \mathcal{S}^2 \to \mathcal{P}^2$ be the mapping that sends $\mathbf{x}$ to $\{\mathbf{x}, -\mathbf{x}\}$. The π is a two-to-one map of $\mathcal{S}^2$ onto $\mathcal{P}^2$.
- A line of $\mathcal{P}^2$ is a set of the form $\pi\mathbf{l}$, where $\mathbf{l}$ is a line of $\mathcal{S}^2$. Clearly, $\pi\mathbf{x}$ lies on $\pi\mathbf{l}$ if and only if $\xi^t\mathbf{x} = 0$.
- **Homogeneous coordinates:** In general, points of real n-dimensional **projective space**, $\mathcal{P}^n$,

  are represented by $n+1$ component column vectors $(x_1, \ldots, x_n, x_{n+1}) \in \mathbb{R}^{n+1}$ such that at least one x_i is non-zero and $(x_1, \ldots, x_n, x_{n+1})$ and $(\lambda x_1, \ldots, \lambda x_n, \lambda x_{n+1})$ represent the same point of $\mathcal{P}^n$ for all $\lambda \neq 0$.
- $(x_1, \ldots, x_n, x_{n+1})$ is the homogeneous representation of a projective point.

4.4 Canonical injection of $\mathbb{R}^n$ into $\mathcal{P}^n$ and Points at infinity

- Affine space $\mathbb{R}^n$ can be embedded in $\mathcal{P}^n$ by

$$(x_1, \ldots, x_n) \to (x_1, \ldots, x_n, 1)$$

- Affine points can be recovered from projective points with $x_{n+1} \neq 0$ by

$$(x_1, \ldots, x_n) \sim (\frac{x_1}{x_{n+1}}, \ldots, \frac{x_n}{x_{n+1}}, 1) \to (\frac{x_1}{x_{n+1}}, \ldots, \frac{x_n}{x_{n+1}})$$

- A projective point with $x_{n+1} = 0$ corresponds to a **point at infinity**.
- The ray $(x_1, \ldots, x_n, 0)$ can be viewed as an additional **ideal point** as $(x_1, \ldots, x_n)$ recedes to infinity in a certain direction. For example, in $\mathcal{P}^2$,

$$\lim_{T \to 0}(X/T, Y/T, 1) = \lim_{T \to 0}(X, Y, T) = (X, Y, 0)$$

It is important to note that these are affine concepts.

4.5 Lines and conics in $\mathcal{P}^2$

- A line equation in $\mathbb{R}^2$ is

$$a_1x_1 + a_2x_2 + a_3 = 0$$

- Substituting by homogeneous coordinates $x_i = X_i/X_3$ we get a homogeneous linear equation

$$(a_1, a_2, a_3) \cdot (X_1, X_2, X_3) = \sum_{i=1}^{3} a_i X_i = 0, \ \mathbf{X} \in \mathcal{P}^2$$

- A line in $\mathcal{P}^2$ is represented by a homogeneous 3-vector (a_1, a_2, a_3).

- A point on a line:

$$\mathbf{a} \cdot \mathbf{X} = 0 \text{ or } \mathbf{a}^T\mathbf{X} = 0 \text{ or } \mathbf{X}^T\mathbf{a} = 0$$

- Two points define a line:

$$\mathbf{l} = \mathbf{p} \times \mathbf{q}$$

- Two lines define a point:

$$\mathbf{x} = \mathbf{l} \times \mathbf{m}$$

- Matrix notation for cross products:
  The cross product $\mathbf{v} \times \mathbf{x}$ can be represented as a matrix multiplication

$$\mathbf{v} \times \mathbf{x} = [\mathbf{v}]_\times \mathbf{x}$$

  where $[\mathbf{v}]_\times$ is a 3×3 antisymmetric matrix of rank 2:

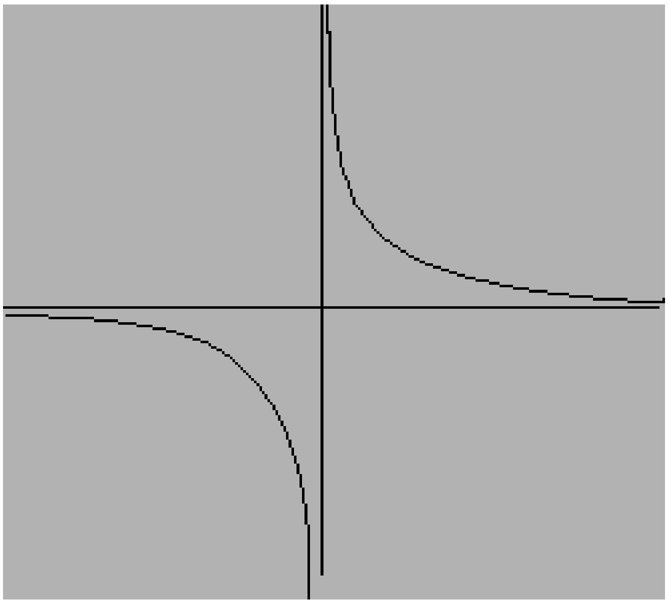
$$[\mathbf{v}]_\times = \begin{bmatrix} 0 & -v_z & v_y \\ v_z & 0 & -v_x \\ -v_y & v_x & 0 \end{bmatrix}$$

- The **line at infinity** ($\mathbf{l}_\infty$): is the line of equation $X_3 = 0$. Thus, the homogeneous representation of $\mathbf{l}_\infty$ is $(0, 0, 1)$.

- The line (u_1, u_2, u_3) intersects $\mathbf{l}_\infty$ at the point $(-u_2, u_1, 0)$.

- Points on $\mathbf{l}_\infty$ are directions of affine lines in the embedded affine space (can be extended to higher dimensions).

- Consider the standard hyperbola in the affine space given by equation $xy = 1$. To transform to homogeneous coordinates, we substitute $x = X/T$ and $y = Y/T$ to obtain $XY = T^2$. This is homogeneous in degree 2. Note that both $(0, \lambda, 0)$ and $(\lambda, 0, 0)$ are solutions. The homogeneous hyperbola crosses the coordinate axes smoothly and emerges from the other side. See the figure.



- A **conic** in affine space (inhomogeneous coordinates) is

$$ax^2 + by^2 + cxy + dx + ey + f = 0$$

Homogenizing this by replacements $x = X_1/X_3$ and $y = Y_1/Y_3$, we obtain

$$aX_1^2 + bX_2^2 + cX_1X_2 + dX_1X_3 + eX_2X_3 + fX_3^2 = 0$$

which can be written in matrix notation as

$$\mathbf{X}^T\mathbf{C}\mathbf{X} = 0$$

where C is symmetric and is the *homogeneous representation* of a **conic**.

4.6 Planes and lines in $\mathcal{P}^3$

The *duality* that exist between points and lines in $\mathcal{P}^2$ exist between points and planes in $\mathcal{P}^3$. Thus a plane is defined as a 4-tuple (u_1, u_2, u_3, u_4) and the equation of this plane is given as

$$\sum_{i=1}^{4} u_i x_i = 0$$

Analogous to the line at infinity ($\mathbf{l}_\infty$) in $\mathcal{P}^2$ we have the *plane at infinity* (π_∞) in $\mathcal{P}^3$ whose representation is $(0, 0, 0, 1)^T$.

4.6.1 Lines is $\mathcal{P}^3$: Pl$\ddot{u}$cker coordinates

4.7 Projective basis

Projective basis: A **projective basis** for $\mathcal{P}^n$ is any set of $n+2$ points no $n+1$ of which are linearly dependent.

Canonical basis:

$$\underbrace{\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \ldots}_{\text{points at infinity along each axis}} \underbrace{\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}}_{\text{origin}}, \underbrace{\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}}_{\text{unit point}}$$

Change of basis: Let $\mathbf{e_1}, \mathbf{e_2}, \ldots, \mathbf{e_{n+1}}, \mathbf{e_{n+2}}$ be the standard basis and $\mathbf{a_1}, \mathbf{a_2}, \ldots, \mathbf{a_{n+1}}, \mathbf{a_{n+2}}$ be any other basis. There exists a non-singular transformation $[\mathbf{T}]_{(n+1)\times(n+1)}$ such that:

$$\mathbf{Te_i} = \lambda_i \mathbf{a_i}, \forall i = 1, 2. \ldots, n+2$$

$\mathbf{T}$ is unique up to a scale.

Proof:
From the first $n+1$ equations we have that $\mathbf{T}$ must be of the form

$$\mathbf{T} = \begin{bmatrix} \lambda_1 \mathbf{a_1} & \lambda_2 \mathbf{a_2} & \ldots & \lambda_{n+1} \mathbf{a_{n+1}} \end{bmatrix}$$

$\mathbf{T}$ is non-singular by the linear independence of $\mathbf{a}$'s.

The final equation gives us:

$$\begin{bmatrix} \lambda_1 \mathbf{a_1} & \lambda_2 \mathbf{a_2} & \ldots & \lambda_{n+1} \mathbf{a_{n+1}} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \lambda_{n+2} \mathbf{a}_{n+2}$$

which is equivalent to:

$$\begin{bmatrix} \mathbf{a_1} & \mathbf{a_2} & \ldots & \mathbf{a_{n+1}} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_{n+1} \end{bmatrix} = \lambda_{n+2} \mathbf{a}_{n+2}$$

Since the matrix on the left hand side of the above equation is of full rank (by linear independence of $\mathbf{x}$'s), the ratios of the λ_i are uniquely determined and no λ_i is 0.

4.8 Collineations

The invertible transformation $\mathbf{T} : \mathcal{P}^n \to \mathcal{P}^n$ is called a **projective transformation** or **collineation** or **homography** or **perspectivity** and is completely determined by $n+2$ point correspondences.
Properties:

4.8.1 Preserves straight lines and cross ratios

Given four collinear points $\mathbf{A}_1$, $\mathbf{A}_2$,$\mathbf{A}_3$ and $\mathbf{A}_4$, their **cross ratio** is defined as

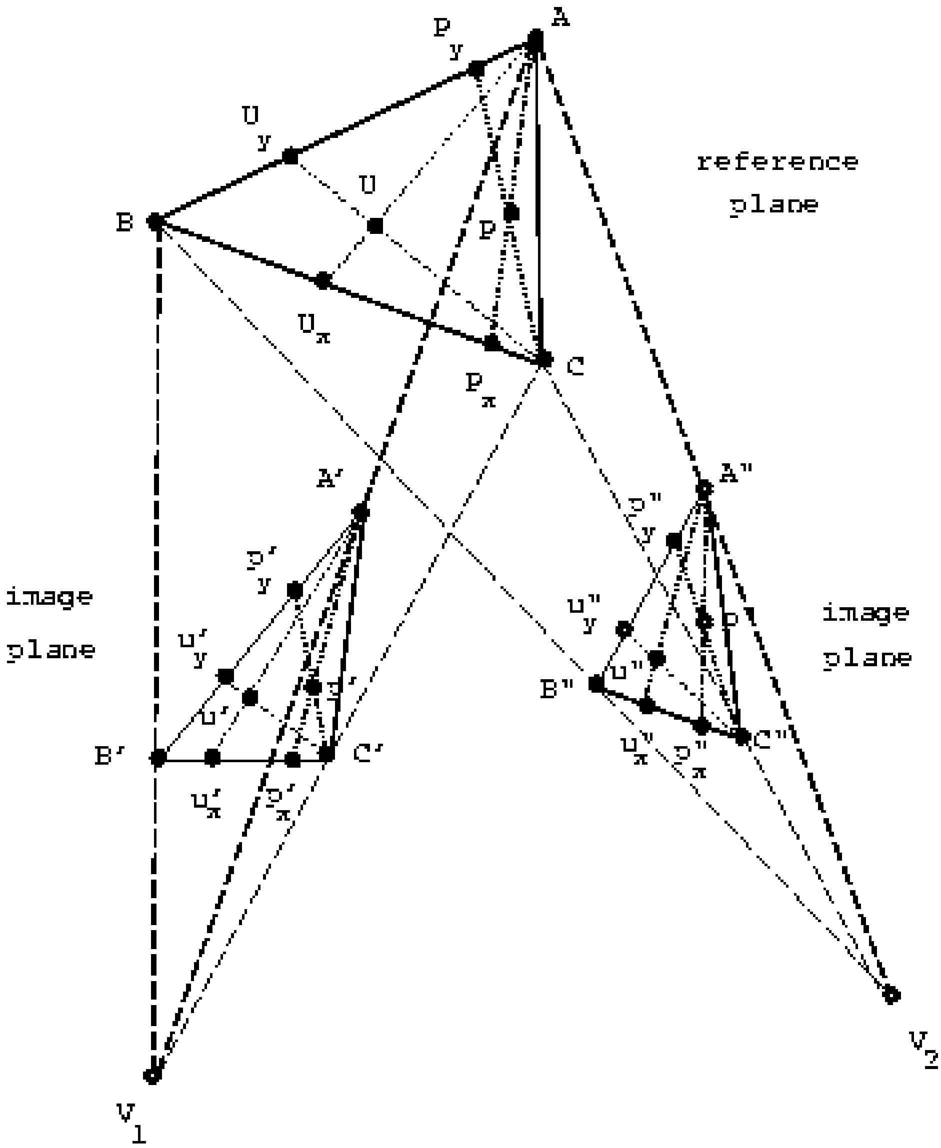
$$\frac{\overline{\mathbf{A}_1\mathbf{A}_3}\;\overline{\mathbf{A}_2\mathbf{A}_4}}{\overline{\mathbf{A}_1\mathbf{A}_4}\;\overline{\mathbf{A}_2\mathbf{A}_3}}$$

if $\mathbf{A}_4$ is a point at infinity the the cross ratio is given as

$$\frac{\overline{\mathbf{A}_1\mathbf{A}_3}}{\overline{\mathbf{A}_2\mathbf{A}_3}}$$

The cross ratio is independent of the choice of the projective coordinate system.

4.8.2 Illustration of perspectivity



4.8.3 Projective mappings of lines and conics in $\mathcal{P}^2$

lines: Let $\mathbf{x}_i$ be a set of points on a line $\mathbf{l}$ and consider the action of a 3×3 projective transformation $\mathbf{H}$ on the the points. Since the points lie on the line we have

$$\mathbf{l}^T \mathbf{x}_i = 0$$

One can easily verify that

$$\mathbf{l}^T \mathbf{H}^{-1} \mathbf{H} \mathbf{x}_i = 0$$

Thus the points $\mathbf{H}\mathbf{x}_i$ all lie on the line $\mathbf{H}^{-T}\mathbf{l}$. Hence, if points are transformed as $\mathbf{x}'_i = \mathbf{H}\mathbf{x}_i$, lines are transformed as $\mathbf{l}' = \mathbf{H}^{-T}\mathbf{l}$.

conics: Note that a conic is represented (homogeneously) as

$$\mathbf{x}^T \mathbf{C} \mathbf{x} = 0$$

Under a point transformation $\mathbf{x}' = \mathbf{H}\mathbf{x}$ the conic becomes

$$\mathbf{x}^T \mathbf{C} \mathbf{x} = \mathbf{x}'^T [\mathbf{H}^{-1}]^T \mathbf{C} \mathbf{H}^{-1} \mathbf{x}' = \mathbf{x}'^T \mathbf{H}^{-T} \mathbf{C} \mathbf{H}^{-1} \mathbf{x}' = 0$$

which is the quadratic form of $\mathbf{x}'^T \mathbf{C}' \mathbf{x}'$ with $\mathbf{C}' = \mathbf{H}^{-T} \mathbf{C} \mathbf{H}^{-1}$. This gives the transformation rule for a conic.

4.8.4 The affine subgroup

In an affine space $\mathcal{A}^n$ an **affine transformation** defines a correspondence $\mathbf{X} \leftrightarrow \mathbf{X}'$ given by:

$$\mathbf{X}' = \mathbf{A}\mathbf{X} + \mathbf{b}$$

where $\mathbf{X}$, $\mathbf{X}'$ and $\mathbf{b}$ are n-vectors, and $\mathbf{A}$ is an $n \times n$ matrix.

Clearly this is a subgroup of the projective group. Its projective representation is

$$\mathbf{T} = \begin{bmatrix} \mathbf{C} & \mathbf{c} \\ \mathbf{0}_n^T & t_{33} \end{bmatrix}$$

where $\mathbf{A} = \frac{1}{t_{33}}\mathbf{C}$ and $\mathbf{b} = \frac{1}{t_{33}}\mathbf{c}$.
The affine subgroup preserves the hyperplane at infinity.

4.8.5 The Euclidean subgroup

- The affine subgroup can be further specialized by the requirement they leave a special conic invariant. The conic $\mathbf{\Omega}_\infty$ is intersection of the quadric of equation:

$$\sum_{i=1}^{n+1} x_i^2 = x_{n+1} = 0 \text{ with } \pi_\infty$$

In a metric frame $\pi_\infty = (0, 0, 0, 1)^T$, and points on $\mathbf{\Omega}_\infty$ satisfy

$$\left. \begin{array}{r} X_1^2 + X_2^2 + X_3^3 \\ X_4 \end{array} \right\} = 0$$

- $\mathbf{\Omega}_\infty$ is called the **absolute conic**. In π_∞, it can be interpreted as a circle of radius $i = \sqrt{-1}$.

- For directions on π_∞ (with $X_4 = 0$), the absolute conic $\mathbf{\Omega}_\infty$ can be expressed as

$$(X_1, X_2, X_3)\mathbf{I}(X_1, X_2, X_3)^T = 0$$

- **The absolute conic, $\mathbf{\Omega}_\infty$, is fixed under a projective transformation H *if and only if* H is an Euclidean transformation**.

  Proof: Since the absolute conic lies on π_∞, a transformation fixing it must also fix π_∞, hence it must be affine. Such a transformation is of the form

$$\mathbf{H}_A = \begin{bmatrix} \mathbf{A} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix}$$

  Restricting to π_∞, the absolute conic is represented by the matrix $\mathbf{I}_{3\times 3}$, and since it is fixed by $\mathbf{H}_A$, one has (up to scale)

$$\mathbf{A}^{-T}\mathbf{I}\mathbf{A}^{-1} = I$$

  which yields $\mathbf{A}\mathbf{A}^T = \mathbf{I}$. This means that $\mathbf{A}$ is orthogonal, hence a scaled rotation. □

4.8.6 How to compute a homography

We are given 2D to 2D point correspondences $\mathbf{x_i} \leftrightarrow \mathbf{x_i'}$ (these are points in $\mathcal{P}^2$ and hence are homogeneous vectors of size 3×1), and we have to find the homography $\mathbf{H}$ (3×3 matrix) such that $\mathbf{x_i'} = \mathbf{H}\mathbf{x_i}$.

Note that $\mathbf{x_i'}$ and $\mathbf{H}\mathbf{x_i}$ are not numerically equal and they can differ by a scale factor. However, they have the same direction, and, hence $\mathbf{x_i'} \times \mathbf{H}\mathbf{x_i} = \mathbf{0}$.

Writing the j^{th} row of $\mathbf{H}$ as $\mathbf{h}^{j^T}$, we have

$$\mathbf{H}\mathbf{x_i} = \begin{bmatrix} \mathbf{h}^{1^T}\mathbf{x_i} \\ \mathbf{h}^{2^T}\mathbf{x_i} \\ \mathbf{h}^{3^T}\mathbf{x_i} \end{bmatrix}$$

Writing $\mathbf{x_i'}^T = (x_i', y_i', w_i')^T$, the cross product becomes:

$$\mathbf{x_i'} \times \mathbf{H}\mathbf{x_i} = \begin{bmatrix} y_i'\mathbf{h}^{3^T}\mathbf{x_i} - w_i'\mathbf{h}^{2^T}\mathbf{x_i} \\ w_i'\mathbf{h}^{1^T}\mathbf{x_i} - x_i'\mathbf{h}^{3^T}\mathbf{x_i} \\ x_i'\mathbf{h}^{2^T}\mathbf{x_i} - y_i'\mathbf{h}^{1^T}\mathbf{x_i} \end{bmatrix}$$

Since $\mathbf{h}^{j^T}\mathbf{x_i} = \mathbf{x_i}^T\mathbf{h}^j$, the system of equations $\mathbf{x_i'} \times \mathbf{H}\mathbf{x_i} = \mathbf{0}$ can be written in terms of the unknowns (the entries of $\mathbf{H}$) as:

$$\begin{bmatrix} \mathbf{0}^T & -w_i'\mathbf{x_i}^T & y_i'\mathbf{x_i}^T \\ w_i'\mathbf{x_i}^T & \mathbf{0}^T & -x_i'\mathbf{x_i}^T \\ -y_i'\mathbf{x_i}^T & x_i'\mathbf{x_i}^T & \mathbf{0}^T \end{bmatrix} \begin{bmatrix} \mathbf{h}^1 \\ \mathbf{h}^2 \\ \mathbf{h}^3 \end{bmatrix} = \mathbf{0}$$

These equations have the form $\mathbf{A_i h} = \mathbf{0}$ where $\mathbf{A_i}$ is a 3×9 matrix and $\mathbf{h}$ is a 9×1 vector (the entries of $\mathbf{H}$). Note that $\mathbf{A_i}$ has rank of 2 (third row is obtained, up to a scale, by a sum of x'_i times the first row and y'_i times the second), and, consequently, for each point correspondence we have really only two equations. We may choose to work with only the first two, but it doesn't harm to keep all three. It may be useful to keep all three equations because if $w'_i = 0$ (a point at infinity), then the first two collapse to a single equation.

Stacking up the equations for $i = 1, 2, 3, 4$ (four points) we have $\mathbf{Ah} = \mathbf{0}$ where $\mathbf{A}$ is a 12×9 matrix whose rank is 8 (of-course, you will not choose four points such that any three are collinear). Consequently $\mathbf{A}$ has a 1-dimensional null space which provides a solution for $\mathbf{h}$. Such a solution can only be determined up to a non-zero scale factor, which suits you fine because $\mathbf{H}$ is anyway defined only up to a scale! A scale may be arbitrarily chosen for $\mathbf{h}$ by insisting that $\|\mathbf{h}\| = 1$.

One can, of-course, stack up more equations by taking more point correspondences. The resulting over-determined system $\mathbf{Ah} = \mathbf{0}$ may not have a solution at all (inconsistent measurements?). We can still find a least-squares solution: *minimize* $\|\mathbf{Ah}\|$ *subject to* $\|\mathbf{h}\| = 1$.

In either case $\mathbf{h}$ is given by the last column of $\mathbf{V}$ where $\mathbf{A} = \mathbf{U\Sigma V}^T$ is the *singular value decomposition (SVD)* of $\mathbf{A}$.

5 Camera models

A Camera transforms a 3D scene point $\mathbf{X} = (X, Y, Z)^T$ into an image point $\mathbf{x} = (x, y)^T$.

The Projective Camera

The most general mapping from $\mathcal{P}^3$ to $\mathcal{P}^2$ is

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} T_{11} & T_{12} & T_{13} & T_{14} \\ T_{21} & T_{22} & T_{23} & T_{24} \\ T_{31} & T_{32} & T_{33} & T_{34} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix}$$

where $(x_1, x_2, x_3)^T$ and $(X_1, X_2, X_3, X_4)^T$ are homogeneous coordinates related to $\mathbf{x}$ and $\mathbf{X}$ by

$$\begin{array}{rcl} (x, y) & = & (x_1/x_3, x_1/x_3) \\ (X, Y, Z) & = & (X_1/X_4, X_2/X_4, X_3/X_4) \end{array}$$

The transformation matrix $\mathbf{T} = [T_{ij}]$ has 11 degrees of freedom since only the ratios of elements T_{ij} are important.

The Perspective Camera

A special case of the projective camera is the **perspective** (or **central**) projection, reducing to the familiar **pin-hole** camera when the leftmost 3×3 sub-matrix of $\mathbf{T}$ is a *rotation matrix* with its third row scaled by the inverse focal length $1/f$. The simplest form is:

$$\mathbf{T}_p = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/f & 0 \end{bmatrix}$$

which gives the familiar equations

$$\begin{bmatrix} x \\ y \end{bmatrix} = \frac{f}{Z} \begin{bmatrix} X \\ Y \end{bmatrix}$$

Each point is scaled by its individual depth, and all projection rays converge to the optic center.

The Affine Camera

The *affine camera* is a special case of the projective camera and is obtained by constraining the matrix $\mathbf{T}$ such that $T_{31} = T_{32} = T_{33} = 0$, thereby reducing the degrees of freedom from 11 to 8:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} T_{11} & T_{12} & T_{13} & T_{14} \\ T_{21} & T_{22} & T_{23} & T_{24} \\ 0 & 0 & 0 & T_{34} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix}$$

In terms of image and scene coordinates, the mapping takes the form

$$\mathbf{x} = \mathbf{M}\mathbf{X} + \mathbf{t}$$

where $\mathbf{M}$ is a general 2×3 matrix with elements $M_{ij} = T_{ij}/T_{34}$ while $\mathbf{t}$ is a general 2-vector representing the image center.

The affine camera preserves parallelism.

The Weak-Perspective Camera

The affine camera becomes a *weak-perspective camera* when the rows of $\mathbf{M}$ form a uniformly scaled rotation matrix. The simplest form is

$$\mathbf{T}_{wp} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & Z_{ave}/f \end{bmatrix}$$

yielding,

$$\mathbf{M}_{wp} = \frac{f}{Z_{ave}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \text{and} \begin{bmatrix} x \\ y \end{bmatrix} = \frac{f}{Z_{ave}} \begin{bmatrix} X \\ Y \end{bmatrix}$$

This is simply the perspective equation with individual point depths Z_i replaced by an average constant depth Z_{ave}

The weak-perspective model is valid when the average variation of the depth of the object (ΔZ) along the line of sight is small compared to the Z_{ave} and the field of view is small. We see this as follows.

Expanding the perspective projection equation using a Taylor series, we obtain

$$\mathbf{x} = \frac{f}{Z_{ave} + \Delta Z}\begin{bmatrix} X \\ Y \end{bmatrix} = \frac{f}{Z_{ave}}\left(1 - \frac{\Delta Z}{Z_{ave}} + \left(\frac{\Delta Z}{Z_{ave}}\right)^2 - \dots\right)\begin{bmatrix} X \\ Y \end{bmatrix}$$

When $|\Delta Z| << Z_{ave}$ only the zero-order term remains giving the weak-perspective projection. The error in image position is then $\mathbf{x}_{err} = \mathbf{x}_p - \mathbf{x}_{wp}$:

$$\mathbf{x}_{err} = -\frac{f}{Z_{ave}}\left(\frac{\Delta Z}{Z_{ave} + \Delta Z}\right)\begin{bmatrix} X \\ Y \end{bmatrix}$$

showing that a small focal length (f), small field of view (X/Z_{ave} and (Y/Z_{ave}) and small depth variation (ΔZ) contribute to the validity of the model.

The orthographic camera

The affine camera reduces to the case of *orthographic (parallel)* projection when $\mathbf{M}$ represents the first two rows of a rotation matrix. The simplest form is

$$\mathbf{T}_{orth} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

yielding,

$$\mathbf{M}_{orth} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \text{and} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} X \\ Y \end{bmatrix}$$

6 Anatomy of a projective camera

6.1 The optical center

- The projective camera $\tilde{\mathbf{P}}$ has a rank 3 whereas it has 4 columns. Clearly, it has a on dimensional right null space. Suppose the null space is generated by the 4-vector $\mathbf{C}$, that is

$$\tilde{\mathbf{P}}\mathbf{C} = \mathbf{0}$$

- **Claim:** $\mathbf{C}$ is the optical center of the camera $\tilde{\mathbf{P}}$.

  Proof: Consider the line containing $\mathbf{C}$ and any other point $\mathbf{A}$ in 3-space. Points on this line can be represented as

$$\mathbf{X}(\lambda) = \lambda\mathbf{A} + (1-\lambda)\mathbf{C}$$

  Under, the mapping $\mathbf{x} = \tilde{\mathbf{P}}\mathbf{X}$, points on this line are projected to

$$\mathbf{x} = \tilde{\mathbf{P}}\mathbf{X}(\lambda) = \lambda\tilde{\mathbf{P}}\mathbf{A} + (1-\lambda)\tilde{\mathbf{P}}\mathbf{C} = \lambda\tilde{\mathbf{P}}\mathbf{A}$$

  since $\tilde{\mathbf{P}}\mathbf{C} = \mathbf{0}$. Since every point on the line are mapped on to the same image point, the line must be a ray through the camera center. It follows that $\mathbf{C}$ is the camera center because for all choices of $\mathbf{A}$ the line passes through the optical center. □

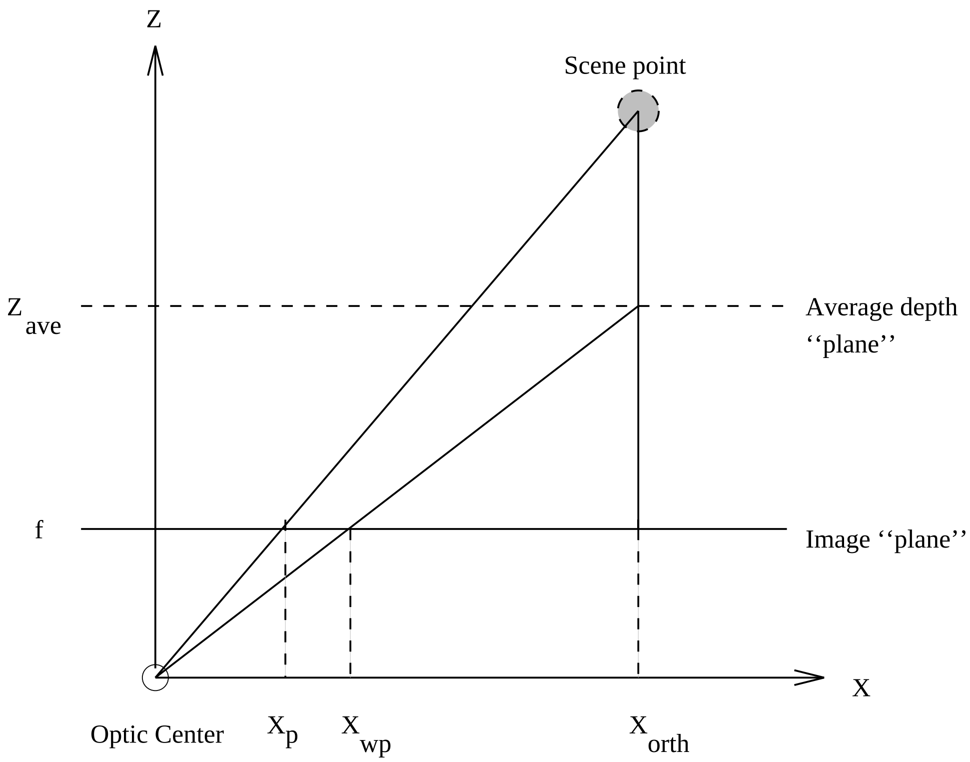


Figure 1: 1D image formation with image plane at $Z = f$. X_p, X_{wp} and X_{orth} are the perspective, weak-perspective and orthographic projections respectively.

- Writing

$$\tilde{\mathbf{P}} = [\mathbf{P} \mid -\mathbf{P}\mathbf{t}]$$

  where $\mathbf{P}$ is 3×3 non-singular we have that $\mathbf{t}$ is the optical center.

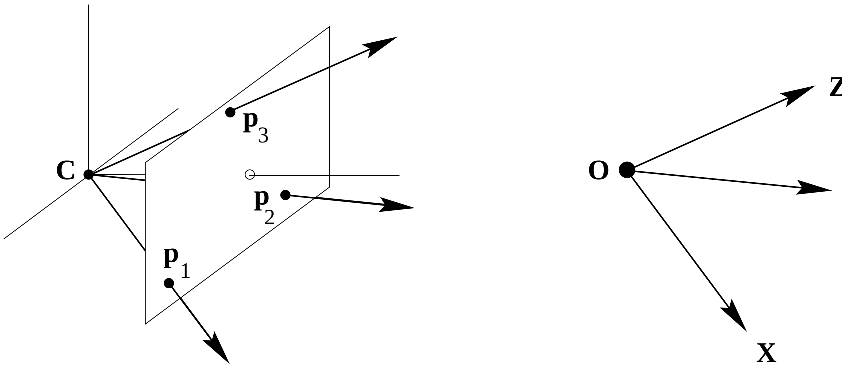
$$[\mathbf{P} \mid -\mathbf{P}\mathbf{t}] \begin{bmatrix} \mathbf{t} \\ 1 \end{bmatrix} = \mathbf{0}$$

- An image point $\mathbf{x}$ defines a line $\lambda \mathbf{P}^{-1}\mathbf{x} + \mathbf{t}$ in 3-space:

$$\mathbf{X} = \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \lambda \begin{bmatrix} \mathbf{P}^{-1} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \\ 0 \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \\ 1 \end{bmatrix}$$

  Clearly all points $\mathbf{X}$ project to $\mathbf{x}$ under $\mathbf{x} = \tilde{\mathbf{P}}\mathbf{X}$.

6.2 The column vectors of the camera matrix



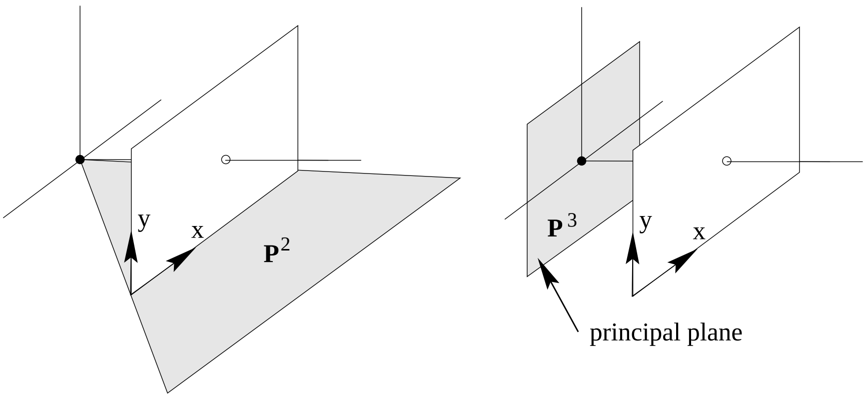
Let the columns of a projective camera $\tilde{\mathbf{P}}$ be $\mathbf{p}_i$ for $i = 1, \ldots, 4$. $\mathbf{p}_i$'s have geometric interpretations as special image points.

$\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$ are the vanishing points of the world coordinates axes X, Y and Z respectively. For example, the X-axis has direction $\mathbf{D} = (1, 0, 0, 0)^T$, which is imaged as

$$\mathbf{p}_1 = \tilde{\mathbf{P}}\mathbf{D}$$

The column $\mathbf{p}_4$ is the image of the world origin $(0, 0, 0, 1)^T$.

6.3 The row vectors of the camera matrix



The row vectors of the projective camera are 4-vectors which have the geometric interpretations as particular world planes. Let us write the rows of the projective camera as π_1^T, π_2^T and π_3^T. That is

$$\tilde{\mathbf{P}} = \begin{bmatrix} \pi_1^T \\ \pi_2^T \\ \pi_3^T \end{bmatrix}$$

6.3.1 The focal plane:

The focal plane is the plane parallel to the image plane containing the optical center. It is the plane of equation

$$\pi_3^T \mathbf{X} = 0$$

Points on this plane project on to image points $(x, y, 0)$, i.e., points at infinity on the image plane.

6.3.2 The axes planes:

Consider points on the plane π_1^T. This set satisfies

$$\pi_1^T \mathbf{X} = 0$$

and, hence, points on this plane project on to image points $(0, y, w)$, which are points on the image y-axis. It also follows from $\tilde{\mathbf{P}}\mathbf{C} = 0$ that $\mathbf{C}$ also lies on π_1^T. Hence π_1^T is the plane defined by the optical center and the y-axis in the image plane.

Similarly, π_2^T is the plane defined by the optical center and the x-axis in the image plane.

Thus unlike π_3^T, π_1^T and π_2^T are dependent on the choice of the coordinate system on the image plane. In particular, the intersection of these two planes is the line joining the optical center with the coordinate origin in the image plane. This line will not, in general, coincide with the principal axis defined below.

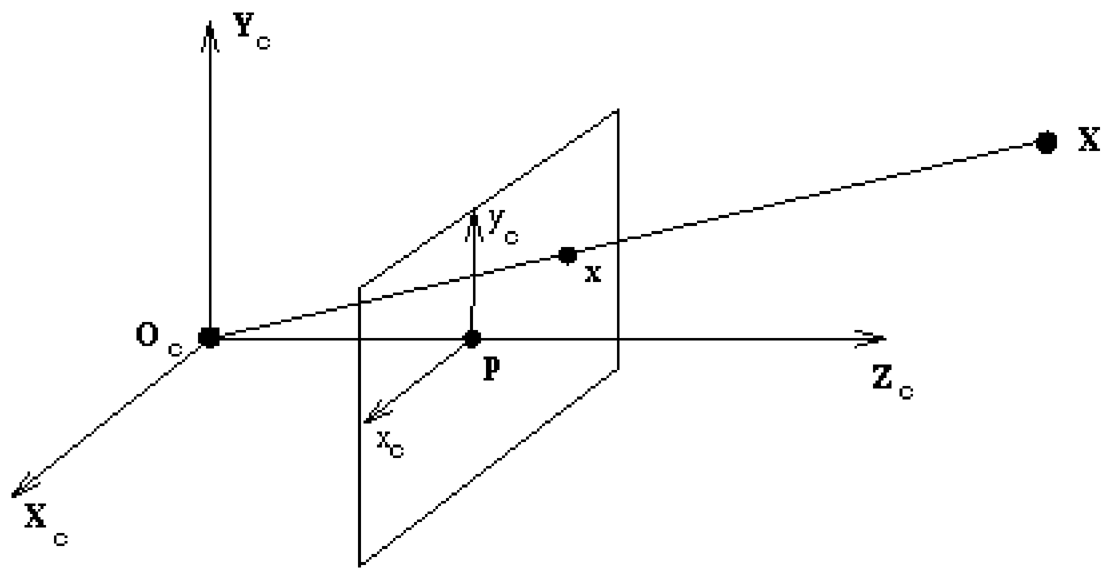
6.4 The Principal axis and the principal point

The *principal axis* is the line passing through the the camera center $\mathbf{C}$ perpendicular to the *focal plane*. It pierces the image plane at the principal point.

6.5 Pin-hole camera revisited

There are three coordinate systems involved - camera, image and the world:

1. **Camera:** perspective projection.

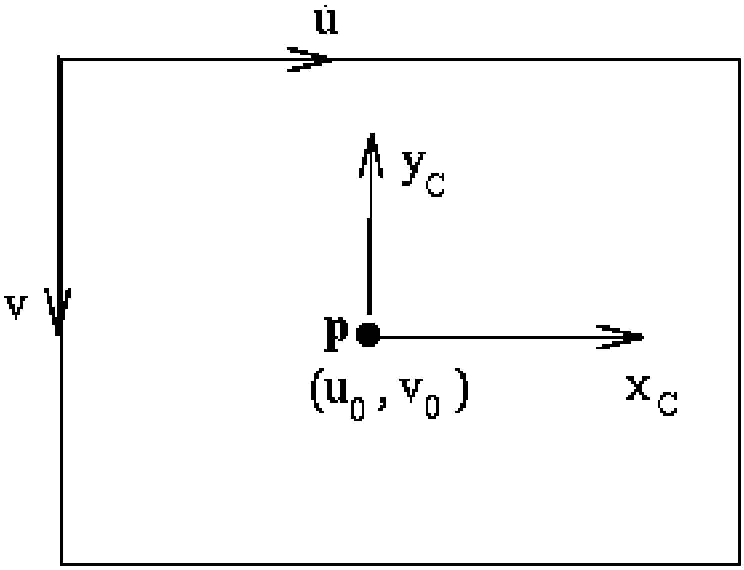


$$\begin{bmatrix} x_c \\ y_c \\ f \end{bmatrix} = k \begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix}$$

where $k = f/Z_c$. This can be written as

$$\begin{bmatrix} x_c \\ y_c \\ f \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix}$$

2. **Image:** (intrinsic/internal camera parameters)



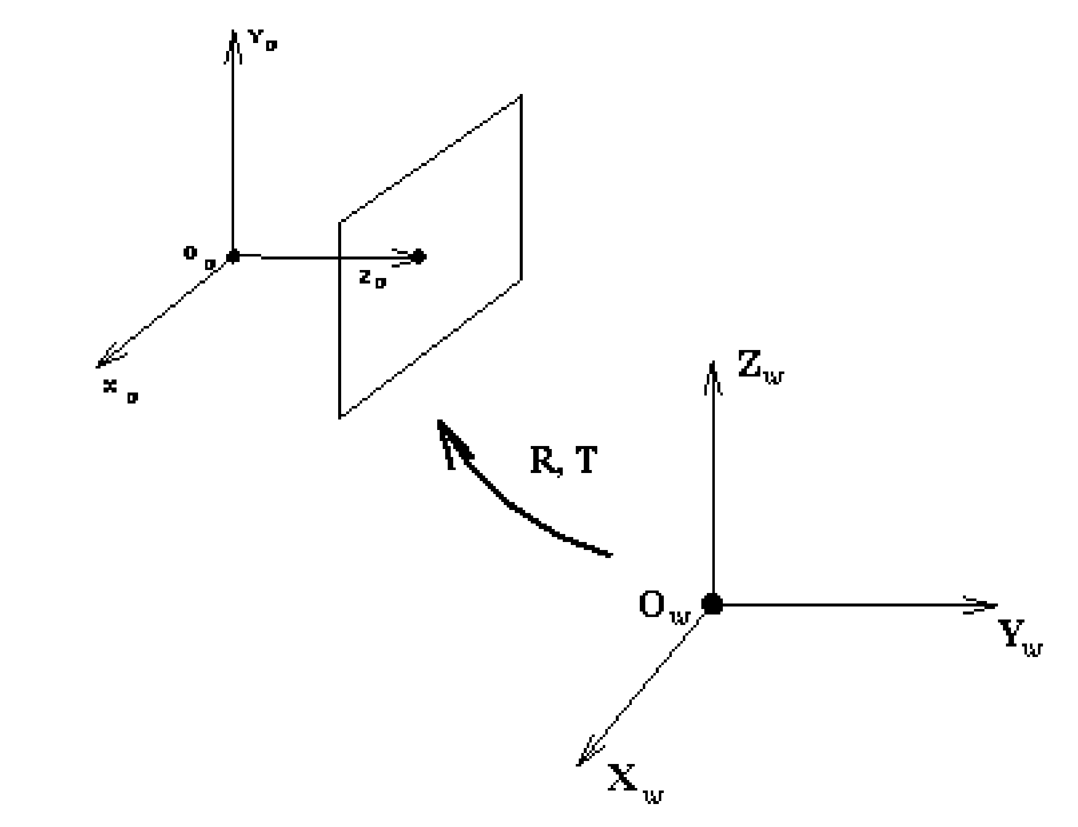
$$\begin{aligned} k_u x_c &= u - u_0 \\ k_v y_c &= v_0 - v \end{aligned}$$

where the unit of k's are pixel/length. This can be expressed as

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} fk_u & 0 & 0 \\ 0 & -fk_v & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ f \end{bmatrix} = \mathbf{C} \begin{bmatrix} x_c \\ y_c \\ f \end{bmatrix}$$

C is called the **camera calibration matrix** and it provides the transformation between an image point and a ray in Euclidean 3-space.

3. **World:** (extrinsic/external camera parameters)



The Euclidean transformation between the camera and world coordinates is:

$$\mathbf{X_c} = \mathbf{R}\mathbf{X_w} + \mathbf{T}$$

and is expressed projectively as:

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R} & \mathbf{T} \\ \mathbf{0}_3^T & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}$$

Finally, concatenating the three matrices, we have

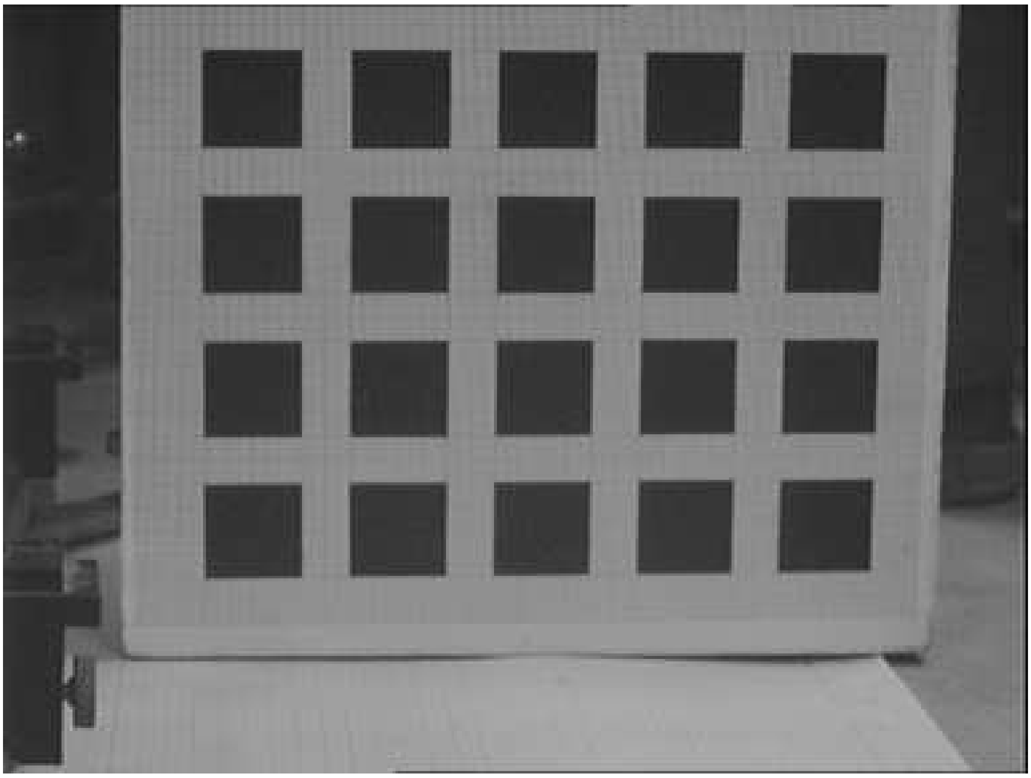
$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{C}\,[\mathbf{R} \mid \mathbf{T}] \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}$$

which defines the 3×4 projection from Euclidean 3-space to an image:

$$\mathbf{x} = \mathbf{P_E}\mathbf{X} \qquad \mathbf{P_E} = \mathbf{C}\,[\mathbf{R} \mid \mathbf{T}]$$

7 Camera calibration

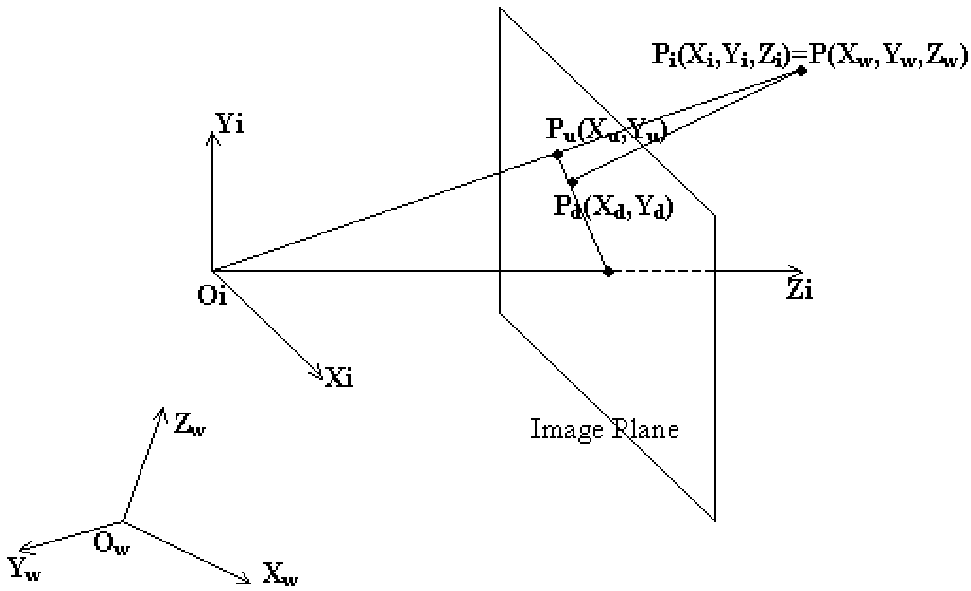
- Classical methods of camera calibration use an calibration object:



- Projection of each point gives us two equations and there are 11 unknowns. 6 points in general position are sufficient for calibration. More points facilitate robust estimation using non-linear least squares.

- Real cameras often have radial distortion.

7.1 Tsai camera model and calibration



- f is the focal length of the camera
- k is the lens radial distortion coefficient
- (u_0, v_0) is the principal point and the center of radial lens distortion
- (R_x, R_y, R_z) are rotation angles for the transformation between the world and camera coordinates
- (t_x, t_y, t_z) are translation components for the transformation between the world and camera coordinates
- $$\begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix} = \mathbf{R} \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} =$$
- $x_i^u = f\frac{X_i}{Z_i}$ and $y_i^u = f\frac{Y_i}{Z_i}$
- $x_i^u = x_i^d(1 + kr^2)$ and $y_i^u = y_i^d(1 + kr^2)$ where $r = \sqrt{(x_i^d)^2 + (y_i^d)^2}$
- Finally, the pixel coordinates are

$$x_i = \frac{s_x x_i^d}{d_x} + u_0 \text{and} y_i = \frac{y_i^d}{d_y} + v_0$$

The parameters of the Tsai model can be estimated either from

1. a coplanar set of world points (a planar calibration object), or
2. a non-coplanar set of world points

In the first case s_x cannot be determined. In what follows we describe non-coplanar calibration, which is a two stage process:

- Linear estimation of a subset of parameters.
- Nonlinear optimization with the linear estimates as initial guess.

7.1.1 Linear estimation of parameters

- Ignoring radial distortion (for the time being) and setting $d_x = d_y = 1$ (measuring f in pixels), we have $x_i^d = x_i^u$ and $y_i^d = y_i^u$.
- Then, combining equations we have

$$\frac{x_i - u_0}{f} = s_x \frac{r_{11}X_w + r_{12}Y_w + r_{13}Z_w + t_x}{r_{31}X_w + r_{32}Y_w + r_{33}Z_w + t_z}$$

  and

$$\frac{y_i - u_0}{f} = \frac{r_{21}X_w + r_{22}Y_w + r_{23}Z_w + t_y}{r_{31}X_w + r_{32}Y_w + r_{33}Z_w + t_z}$$

- Assuming (u_0, v_0) to be known (at the center of the image) and setting $x_i' = x_i - u_0$ and $y_i' = y_i - v_0$), we have

$$\frac{x_i'}{f} = s_x \frac{X_i}{Z_i} \text{ and } \frac{y_i'}{f} = \frac{Y_i}{Z_i}$$

- Eliminating f we have

$$\frac{x_i'}{y_i'} = s_x \frac{X_i}{Y_i} = s_x \frac{r_{11}X_w + r_{12}Y_w + r_{13}Z_w + t_x}{r_{21}X_w + r_{22}Y_w + r_{23}Z_w + t_y}$$

- Rearranging, we have

$$s_x \left(r_{11}X_w + r_{12}Y_w + r_{13}Z_w + t_x\right) y_i' - \left(r_{21}X_w + r_{22}Y_w + r_{23}Z_w + t_y\right) x_i' = 0$$

  or

$$(X_w y_i')s_x r_{11} + (Y_w y_i')s_x r_{12} + (Z_w y_i')s_x r_{13} + y_i' s_x t_x - (X_w x_i')r_{21} + (Y_w x_i')r_{22} + (Z_w x_i')r_{23} + x_i' t_y = 0$$

  which is a linear homogeneous equation in the eight unknowns $s_x r_{11}$, $s_x r_{12}$, $s_x r_{13}$, r_{21}, r_{22}, r_{23}, $s_x t_x$ and t_y.

- The unknown scale factor can be fixed by setting $t_y = 1$. Image correspondences of seven points in general position are sufficient to solve for the remaining unknowns. Let the solution be $s_x r_{11}'$, $s_x r_{12}'$, $s_x r_{13}'$, r_{21}', r_{22}', r_{23}', $s_x t_x'$ and $t_y' = 1$
- We can estimate the correct scale factor by noting that the two rows of the rotation matrix are supposed to be normal, i.e.,

$$r_{11}^2 + r_{12}^2 + r_{13}^2 = r_{21}^2 + r_{22}^2 + r_{23}^2 = 1$$

- The scale factor c for the solution can then be determined from

$$c = 1/\sqrt{(r'_{21})^2 + (r'_{22})^2 + (r'_{23})^2}$$

  and

$$c/s_x = 1/\sqrt{(s_x r'_{11})^2 + (s_x r'_{12})^2 + (s_x r'_{13})^2}$$

  This also allows recovery of s_x.

- In the above procedure we didn't enforce orthogonality of the first two rows of R. Given vectors $\mathbf{r'_1}$ and $\mathbf{r'_2}$, we can find two orthogonal vectors $\mathbf{r_1}$ and $\mathbf{r_2}$ close to the originals as follows:

$$\mathbf{r_1} = \mathbf{r'_1} + \mu\mathbf{r'_2} \text{ and } \mathbf{r_2} = \mathbf{r'_2} + \mu\mathbf{r'_1}$$

  which gives

$$\mathbf{r_1} \cdot \mathbf{r_2} = \mathbf{r'_1} \cdot \mathbf{r'_2} + \mu(\mathbf{r'_1} \cdot \mathbf{r'_1} + \mathbf{r'_2} \cdot \mathbf{r'_2}) + \mu^2 \mathbf{r'_1} \cdot \mathbf{r'_2} = 0$$

  The solution of this quadratic in μ is numerically ill behaved because $\mathbf{r'_1} \cdot \mathbf{r'_2}$ will be quite small. We can use the approximate solution

$$\mu \simeq -(1/2)\mathbf{r'_1} \cdot \mathbf{r'_2}$$

  since $\mathbf{r'_1} \cdot \mathbf{r'_1}$ and $\mathbf{r'_2} \cdot \mathbf{r'_2}$ are both near 1.

- $\mathbf{r_3}$ can then be recovered as $\mathbf{r_1} \times \mathbf{r_2}$.

- Once we have R we can estimate f and t_z from the basic equations above. This will require one more correspondence to be given.

- The above procedure may be problematic if t_y is close to 0. In such a case the entire experimental data may have to be first translated by a fixed amount.

7.1.2 Nonlinear optimization

Finally, using the linear estimates of R, T and f as a starting point one can solve for all the parameters, including the radial lens distortion parameter k which was initialized to 0, by minimizing the image distance

$$\sum_{i=1}^{N}(x_i - x_i^p)^2 + \sum_{i=1}^{N}(y_i - y_i^p)^2$$

where (x_i, y_i) are the observed image points and (x_i^p, y_i^p) are the positions predicted by the Tsai model. The nonlinear optimization (over all the parameters) can be carried out by an iterative numerical technique like the *Levenberg-Marquardt method*.

7.2 Camera calibration and absolute conic

- Let $\mathcal{C}$ and $\mathcal{R}$ be the optical center and the retinal plane of a camera.
- Consider a rigid motion $\mathcal{D}$ of the camera from configuration $(\mathcal{C}, \mathcal{R})$ to $(\mathcal{D}(\mathcal{C}), \mathcal{D}(\mathcal{R}))$.
- Let ω_1 and ω_2 be the two images of Ω corresponding to the two configurations.
- Clearly $\omega_1 = \omega_2$ because $\Omega = \mathcal{D}(\Omega)$.
- Thus ω is determined by only the internal parameters.
- Consider the equation of Ω

$$X^2 + Y^2 + Z^2 = 0 = T = \mathbf{M}^T\mathbf{M}$$

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

- Images $\mathbf{x}$ of points $\mathbf{X}$ of Ω satisfy the equation

$$\mathbf{x} = [\mathbf{P} \mid -\mathbf{Pt}] \begin{bmatrix} X \\ Y \\ Z \\ 0 \end{bmatrix} = \mathbf{PM}$$

- Since $\mathbf{M}^T\mathbf{M} = 0$, we have that the equation of ω is

$$\mathbf{x}^T\mathbf{P}^{-1T}\mathbf{P}^{-1}\mathbf{x} = 0$$

- In terms of pixel coordinates this can be re-written as:

$$(\frac{u - u_0}{fk_u})^2 + (\frac{v - v_0}{fk_v})^2 + 1 = 0$$

  Thus, the image of the absolute conic are determined completely by the internal parameters.

7.3 What does calibration give?

- An image point $\mathbf{x}$ back projects to a ray defined by $\mathbf{x}$ and the camera center. **Calibration relates the image point to the ray's direction**.
- Suppose points on the ray are written as $\tilde{\mathbf{X}} = \lambda\mathbf{d}$ in the camera Euclidean frame. Then these points map to the point

$$\mathbf{x} = \mathbf{C}\,[\mathbf{I} \mid \mathbf{0}]\,(\lambda\mathbf{d}^T, ?)^T$$

- Thus, $\mathbf{C}$ is the (affine) transformation between $\mathbf{x}$ and the ray's direction $\mathbf{d} = \mathbf{C}^{-1}\mathbf{x}$ measured in the cameras Euclidean frame.

- The angle between two rays $\mathbf{d}_1$ and $\mathbf{d}_2$ corresponding to image points $\mathbf{x}_1$ and $\mathbf{x}_2$ may be obtained as (by the cosine formula)

$$\cos\theta = \frac{\mathbf{d}_1{}^T\mathbf{d}_2}{\sqrt{\mathbf{d}_1{}^T\mathbf{d}_1}\sqrt{\mathbf{d}_2{}^T\mathbf{d}_2}} = \frac{(\mathbf{C}^{-1}\mathbf{x}_1)^T(\mathbf{C}^{-1}\mathbf{x}_2)}{\sqrt{(\mathbf{C}^{-1}\mathbf{x}_1)^T(\mathbf{C}^{-1}\mathbf{x}_1)}\sqrt{(\mathbf{C}^{-1}\mathbf{x}_2)^T(\mathbf{C}^{-1}\mathbf{x}_2)}}$$
$$= \frac{\mathbf{x}_1{}^T(\mathbf{C}^{-T}\mathbf{C}^{-1})\mathbf{x}_2}{\sqrt{\mathbf{x}_1{}^T(\mathbf{C}^{-T}\mathbf{C}^{-1})\mathbf{x}_1}\sqrt{\mathbf{x}_2{}^T(\mathbf{C}^{-T}\mathbf{C}^{-1})\mathbf{x}_2}}$$

- The above shows that if $\mathbf{C}$ is known (*camera is calibrated*), then the angle between rays can be computed from their corresponding image points. *A calibrated camera is like a 2D protractor.*

- An image line $\mathbf{l}$ defines a plane through the camera center with normal direction $\mathbf{n} = \mathbf{C}^T\mathbf{l}$.

  Proof: Points $\mathbf{x}$ on $\mathbf{l}$ back projects to directions $\mathbf{d} = \mathbf{C}^{-1}\mathbf{x}$ which are orthogonal to the plane normal. Hence, $\mathbf{d}^T\mathbf{n} = \mathbf{x}^T\mathbf{C}^{-T}\mathbf{n} = 0$. Since points on $\mathbf{l}$ satisfy $\mathbf{x}^T\mathbf{l} = 0$ we have that $\mathbf{l} = \mathbf{C}^{-T}\mathbf{n}$. □

7.4 The image of the absolute conic

- Points on the **plane at infinity** (π_∞), which may be written as $\mathbf{X}_\infty = (\mathbf{d}^T, 0)^T$ are mapped to the image plane by a general camera $\mathbf{P} = \mathbf{CR}\,[\mathbf{I} \mid \mathbf{t}]$ as

$$\mathbf{x} = \mathbf{PX}_\infty = \mathbf{CR}\,[\mathbf{I} \mid \mathbf{t}]\,(\mathbf{d}^T, 0)^T = \mathbf{CRd}$$

- Thus $\mathbf{H} = \mathbf{CR}$ is the planar homography between π_∞ and the image plane. Note that the mapping is independent of the position (translation) of the camera and depends only on the orientation. (An explanation as to why the images of stars stay fixed on the retinae as we translate?)

- Since the **absolute conic** ($\mathbf{\Omega}_\infty$) is on π_∞, we can compute its image as

$$\omega = (\mathbf{CC}^T)^{-1} = \mathbf{C}^{-T}\mathbf{C}^{-1}$$

  Proof: Note that under a point homography $\mathbf{H}$ which maps $\mathbf{x}$ to $\mathbf{Hx}$, a conic $\mathbf{A}$ is mapped to $\mathbf{H}^{-T}\mathbf{AH}^{-1}$. Hence $\mathbf{\Omega}_\infty = \mathbf{I}$ on π_∞ maps to

$$\omega = (\mathbf{CR})^{-T}\mathbf{I}(\mathbf{CR})^{-1} = \mathbf{C}^{-T}\mathbf{RR}^{-1}\mathbf{C}^{-1} = (\mathbf{CC}^T)^{-1} = \mathbf{C}^{-T}\mathbf{C}^{-1}$$

□

- Like $\mathbf{\Omega}_\infty$, ω is an imaginary point conic with no real points. It cannot really be observed in an image. It is really an useful mathematical device.

- ω depends only on the internal parameters of the camera and is independent of the cameras position or orientation.

- It follows from above that the angle between two rays is given by the simple equation

$$\cos\theta = \frac{\mathbf{x_1}^T(\mathbf{C}^{-T}\mathbf{C}^{-1})\mathbf{x_2}}{\sqrt{\mathbf{x_1}^T(\mathbf{C}^{-T}\mathbf{C}^{-1})\mathbf{x_1}}\sqrt{\mathbf{x_2}^T(\mathbf{C}^{-T}\mathbf{C}^{-1})\mathbf{x_2}}} = \frac{\mathbf{x_1}^T\omega\mathbf{x_2}}{\sqrt{\mathbf{x_1}^T\omega\mathbf{x_1}}\sqrt{\mathbf{x_2}^T\omega\mathbf{x_2}}}$$

- The above expression is independent of the choice of the projective coordinate system on the image. To see this consider any 2D projective transformation $\mathbf{H}$. The points $\mathbf{x}_i$ are transformed to $\mathbf{H}\mathbf{x}_i$, and ω transforms (as any image conic) to $\mathbf{H}^{-T}\omega\mathbf{H}^{-1}$. Hence the expression for $\cos(\theta)$ is unchanged.

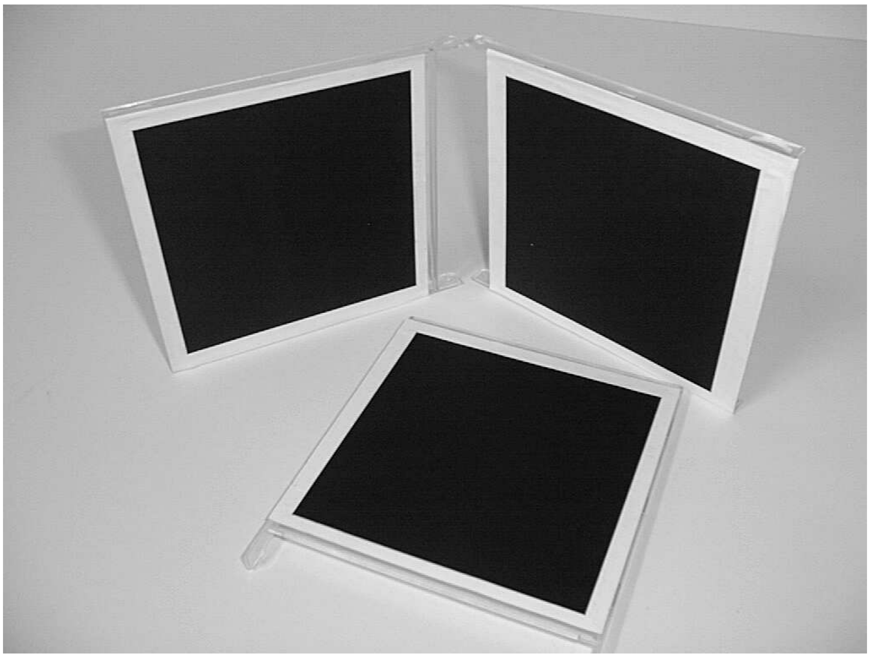
- We may define the dual image of the absolute conic as

$$\omega^* = \omega^{-1} = \mathbf{C}\mathbf{C}^T$$

- Once ω (equivalently ω^*) is identified in an image $\mathbf{C}$ is uniquely determined; since a symmetric matrix can be uniquely decomposed into an upper triangular matrix and its transpose ($\omega^* = \mathbf{C}\mathbf{C}^T$) by Cholesky decomposition.

- An arbitrary plane π intersects π_∞ in a line, and this line intersects Ω_∞ in two points (imaginary) which are *circular points* of π. The image of the circular points line on ω at the points at which the vanishing lines of the plane π intersects ω.

7.5 A simple calibration device



The last two points above can be used to design a simple calibration device as follows. The image of three squares on three different planes (not necessarily orthogonal) are sufficient to give calibration. Consider the following steps:

1. For each square compute the homography $\mathbf{H}$ that maps its corner points, $(0,0)^T, (0,1)^T, (1,0)^T, (1,1)^T$ to their imaged points.

2. Compute the imaged circular points for the plane of that square as $\mathbf{H}(1, \pm i, 0)^T$.

3. Fit a conic ω through the six imaged points. Note that five points are sufficient to define a conic.

4. Compute $\mathbf{C}$ from $\omega = (\mathbf{C}\mathbf{C}^T)^{-1}$ using Cholesky decomposition.

7.6 Vanishing points and vanishing lines

- Points on a line in 3 space through a point $\mathbf{A}$ and with direction $\mathbf{D} = (\mathbf{d}, 0)^T$ can be written as $\mathbf{X}(\lambda) = \mathbf{A} + \lambda\mathbf{D}$. As λ varies from 0 to ∞, $\mathbf{X}(\lambda)$ varies from $\mathbf{A}$ to the the point at infinity $\mathbf{D}$.

- Under a projective camera $\mathbf{P} = \mathbf{C}\,[\mathbf{I} \mid \mathbf{0}]$ the points project as

$$\mathbf{x}(\lambda) = \mathbf{P}\mathbf{X}(\lambda) = \mathbf{P}\mathbf{A} + \lambda\mathbf{P}\mathbf{D} = \mathbf{a} + \lambda\mathbf{C}\mathbf{d}$$

  where $\mathbf{a}$ is the image of $\mathbf{A}$.

- The **vanishing point** of the line, $\mathbf{v}$, is obtained as

$$\mathbf{v} = \lim_{\lambda\to\infty} \mathbf{x}(\lambda) = \lim_{\lambda\to\infty} \mathbf{a} + \lambda\mathbf{C}\mathbf{d} = \mathbf{C}\mathbf{d}$$

  or

$$\mathbf{v} = \mathbf{P}\mathbf{X}_\infty = \mathbf{C}\,[\mathbf{I} \mid \mathbf{0}]\,(\mathbf{d}^T, 0)^T = \mathbf{C}\mathbf{d}$$

  Thus the vanishing point back projects to a ray with direction $\mathbf{d}$.

- Parallel planes in 3 space intersects π_∞ in a common line, and the image of this line is the vanishing line of the plane. Thus, we have
  The set of planes perpendicular to the direction $\mathbf{n}$ *in the camera's Euclidean frame have vanishing line* $\mathbf{l} = \mathbf{C}^{-T}\mathbf{n}$.

7.6.1 Camera rotation from vanishing points

- Consider two images of a scene obtained by the same camera from different position and orientation.

- The images of the points at infinity, **the vanishing points**, are not affected by the camera translation, but are affected only by the camera rotation $\mathbf{R}$.

- Consider a scene line with vanishing point $\mathbf{v}_i$ in the first view and $\mathbf{v}'_i$ in the second.

- The vanishing point $\mathbf{v}_i$ has a direction $\mathbf{d}_i$ in the first cameras Euclidean frame, and, similarly, the vanishing point $\mathbf{v}'_i$ has a direction $\mathbf{d}'_i$ in the second cameras Euclidean frame. We have

$$\mathbf{d}_i = \mathbf{C}^{-1}\mathbf{v}_i / \|\mathbf{C}^{-1}\mathbf{v}_i\|$$
$$\mathbf{d}'_i = \mathbf{C}^{-1}\mathbf{v}'_i / \|\mathbf{C}^{-1}\mathbf{v}'_i\|$$

- The directions are related by

$$\mathbf{d}'_i = \mathbf{R}\mathbf{d}_i$$

  which represents two independent constraints on $\mathbf{R}$.

- Hence, the rotation matrix can be computed from two such corresponding directions provided we know $\mathbf{C}$.

7.6.2 Determining calibration from vanishing points and lines

- Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be the vanishing points of two lines in the image. If θ is the angle between the two scene lines, we have

$$\cos\theta = \frac{\mathbf{v}_1{}^T\omega\mathbf{v}_2}{\sqrt{\mathbf{v}_1{}^T\omega\mathbf{v}_1}\sqrt{\mathbf{v}_2{}^T\omega\mathbf{v}_2}}$$

- If θ is known the above equation gives a quadratic constraint on the entries of ω.

- If it is known that the scene lines are orthogonal ($\theta = 90$), then we have a linear constraint

$$\mathbf{v}_1{}^T\omega\mathbf{v}_2 = 0$$

  Thus, given five pairs of perpendicular lines, one can solve for the entries of ω.

- The vanishing point $\mathbf{v}$ of the normal direction to a plane is obtained from the plane vanishing line as

$$\mathbf{l} = \mathbf{C}^{-T}\mathbf{n} = \mathbf{C}^{-T}\mathbf{C}^{-1}\mathbf{v} = \omega\mathbf{v}$$

  A common example is a vertical direction and a horizontal plane.

- Writing the above as $\mathbf{l} \times \omega\mathbf{v} = \mathbf{0}$ removes the homogeneous scaling factor and results in three homogeneous equations linear in the entries of ω.

- Given a sufficient number of such constraints ω can be computed and $\mathbf{C}$ follows.

- The following can be verified by direct computation:

  1. If $s = \mathbf{C}_{12} = 0$ (no skew) then $\omega_{12} = \omega_{21} = 0$.
  2. If, in addition, $\alpha_x = \alpha_y = \mathbf{C}_{11} = \mathbf{C}_{22}$ then $\omega_{11} = \omega_{22}$

- Suppose it is known that the camera has zero skew and that the pixels are square (or the aspect ratio is known) the ω and $\mathbf{C}$ can be computed from an orthogonal triad of directions.

7.7 Zhang's camera calibration

- Camera calibration from a single plane at few (at least three, two skew is ignored) orientations.
- Without loss of generality, assume that the model plane is on $Z = 0$.
- Then, for points on the model plane

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{C} \begin{bmatrix} \mathbf{r_1} & \mathbf{r_2} & \mathbf{r_3} & \mathbf{t} \end{bmatrix} \begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix} = \mathbf{C} \begin{bmatrix} \mathbf{r_1} & \mathbf{r_2} & \mathbf{t} \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}$$

- Thus, a model point $\mathbf{M}$ and its image $\mathbf{m}$ are related by a homography $\mathbf{H}$, where

$$s\mathbf{m} = \mathbf{HM}$$

  with

$$\mathbf{H} = \mathbf{C} \begin{bmatrix} \mathbf{r_1} & \mathbf{r_2} & \mathbf{t} \end{bmatrix}$$

- From

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \mathbf{h}_3 \end{bmatrix} = \lambda \mathbf{C} \begin{bmatrix} \mathbf{r_1} & \mathbf{r_2} & \mathbf{t} \end{bmatrix}$$

  using the fact that $\mathbf{r_1}$ and $\mathbf{r_2}$ are orthonormal, we obtain the relationships

$$\begin{aligned} \mathbf{h}_1^t \mathbf{C}^{-t} \mathbf{C}^{-1} \mathbf{h}_2 &= 0 \\ \mathbf{h}_1^t \mathbf{C}^{-t} \mathbf{C}^{-1} \mathbf{h}_1 &= \mathbf{h}_2^t \mathbf{C}^{-t} \mathbf{C}^{-1} \mathbf{h}_2 \end{aligned}$$

- Each such homography provides two constraints on the camera intrinsics (image of the absolute conic). Three independent orientations are sufficient to solve for camera internals linearly. Two are sufficient if the skew is ignored.
- Once the camera internals matrix $\mathbf{C}$ is known, the externals can be readily obtained.

$$\begin{aligned} \mathbf{r}_1 &= \lambda \mathbf{C}^{-1} \mathbf{h}_1 \\ \mathbf{r}_2 &= \lambda \mathbf{C}^{-1} \mathbf{h}_2 \\ \mathbf{r}_3 &= \mathbf{r}_1 \times \mathbf{r}_2 \\ \mathbf{t} &= \lambda \mathbf{C}^{-1} \mathbf{h}_3 \end{aligned}$$

  with $\lambda = 1/||\mathbf{C}^{-1}\mathbf{h}_1|| = 1/||\mathbf{C}^{-1}\mathbf{h}_2||$. Of course, the computed matrix $\mathbf{R} = [\mathbf{r}_1 \ \ \mathbf{r}_2 \ \ \mathbf{r}_3]$ does not, in general, satisfy the properties of a rotation matrix. The orthonormality properties can be enforced in manned similar to the one described in Tsai's.
- The method can be extended to also obtain the radial lens distortion parameters.