Notes on Multiple View Geometry in Computer Vision

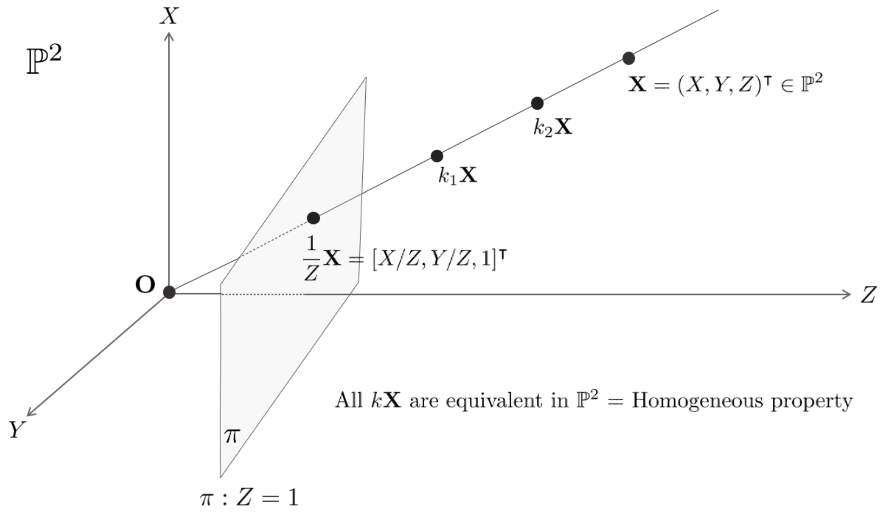
Gyubeom Edward Im*
(Orig. by Richard Hartley and Andrew Zisserman)

May 4, 2024

Contents

*blog: alida.tistory.com, email: criterion.im@gmail.com

1 Projective Space



The projective space $\mathbb{P}^n$ is defined as the set of lines passing through the origin in $\mathbb{R}^{n+1}$ space. Thus, it includes every element of $\mathbb{R}^{n+1}$ except the origin. Strictly speaking, since it only deals with real numbers excluding imaginary numbers, it should be written as $\mathbb{RP}^n$, but for convenience in this posting, we will use $\mathbb{P}^n$.

$$\mathbb{P}^n = \mathbb{R}^{n+1} - \{0\} \tag{1}$$

Let's say a point $\mathbf{X}$ in 3-dimensional space is given as follows.

$$\mathbf{X} = [X, Y, Z] \in \mathbb{P}^2 \tag{2}$$

Even if every element of X is multiplied by an arbitrary value k, it still exists on the line connecting the origin and X, and this property is called the homogeneous property. If $k = 1/Z$ is multiplied, it geometrically means the same as projecting the 3-dimensional point onto the plane where $Z = 1$.

$$[X, Y, Z] \to [X/Z, Y/Z, 1] \tag{3}$$

Therefore, by using $\mathbb{P}^2$, points in 3-dimensional space can be projected onto a specific plane and expressed in 2-dimensional space like $\mathbb{R}^2$, and it also has the expressive advantage of representing additional elements such as the point at infinity $\mathbf{x}_\infty$ and the line at infinity $\mathbf{l}_\infty$. Furthermore, it has the operational advantage of computing points and lines with the same 3-dimensional vector. Details will be explained in the following section.

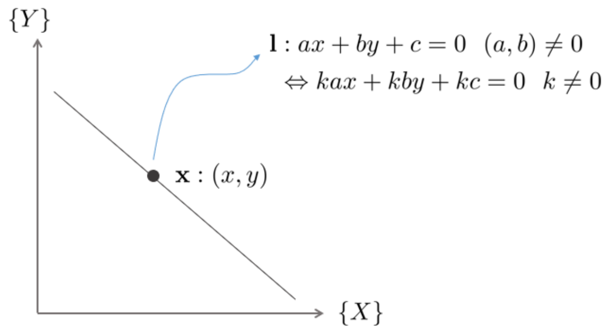
2 Projective Geometry and Transformations in 2D

The 2D projective plane

Typically, a point $\mathbf{x}$ on the plane is expressed as $(x, y) \in \mathbb{R}^2$. If $\mathbb{R}^2$ is a vector space, $\mathbf{x}$ can be represented as a single vector. Moreover, a line $\mathbf{l}$ that includes two points $\mathbf{x}_1, \mathbf{x}_2$ can be represented by subtracting the two vectors. This section explains Homogeneous Notation, which allows points and lines on the plane to be expressed by the same vector.

Points and lines

Homogeneous representation of line



A line $\mathbf{l}$ can be represented as follows:

$$\mathbf{l} : ax + by + c = 0 \quad (a, b) \neq 0 \tag{4}$$

If a point $\mathbf{x} = (x, y, 1)$ exists on the line $\mathbf{l}$, according to the formula $ax + by + c = 0$, the line $\mathbf{l}$ can be expressed as follows:

$$\mathbf{l} : (a, b, c) \tag{5}$$

In this case, (a, b, c) does not uniquely represent line $\mathbf{l}$. Even if a non-zero arbitrary constant k is multiplied, such as (ka, kb, kc), it can still represent the same line $\mathbf{l}$.

$$\mathbf{l} : (ka, kb, kc) \tag{6}$$

Therefore, **the line l on the plane represents the same line regardless of scale value**. All vectors in this equivalent relationship are called Homogeneous vectors. **The set of all vectors in the $\mathbb{R}^3$ space that are in this equivalence relationship is called the projective space $\mathbb{P}^2$.**

Homogeneous representation of points

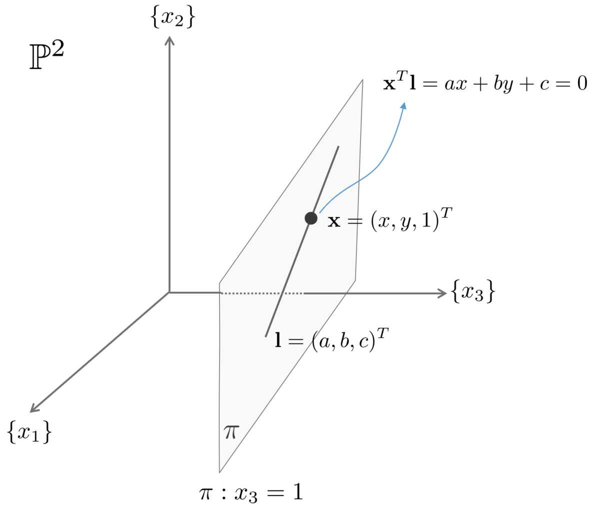
There is the following formula between the line $\mathbf{l} = (a, b, c)^\intercal$ and a point $\mathbf{x} = (x, y)^\intercal$ on the line:

$$ax + by + c = 0 \tag{7}$$

This can be expressed as the dot product of two vectors $\mathbf{l}$ and $\mathbf{x}$.

$$\begin{pmatrix} x & y & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} x & y & 1 \end{pmatrix} \mathbf{l} = 0 \tag{8}$$

This can be seen as adding 1 at the end of the coordinates of a point $\mathbf{x} = (x, y)$ and dot producting it with the line. Since line $\mathbf{l}$ can represent a single line regardless of scale, under the premise that $(x, y, 1)\mathbf{l} = 0$ holds, for all values of k, $(kx, ky, k)\mathbf{l} = 0$ also holds. Therefore, **for any arbitrary constant k, (kx, ky, k) represents one point $\mathbf{x} = (x, y)$ in $\mathbb{R}^2$ space, hence a point can also be represented as a Homogeneous vector like a line**. Generalizing this representation, an arbitrary point $\mathbf{x} = (x, y, z)^\intercal$ represents the point $(x/z, y/z)$ in $\mathbb{R}^2$ space.



Therefore, if an arbitrary point $\mathbf{x}$ exists on the line $\mathbf{l}$ in the space $\mathbb{P}^2$,

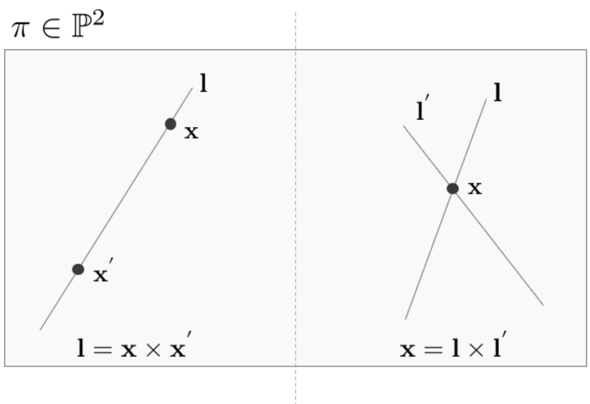
$$
\begin{aligned}
\mathbf{x}^\intercal \mathbf{l} &= \begin{bmatrix} x & y & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} \\
&= ax + by + c \\
&= 0 \\
&\therefore \mathbf{x}^\intercal \mathbf{l} = 0
\end{aligned}
\tag{9}
$$

The above formula is established.

Degrees of freedom (dof)

In the space $\mathbb{P}^2$, a single point is uniquely determined by two values (x, y). Similarly, to uniquely determine a line, two independent ratios $\{a : b : c\}$ must be provided. Thus, in the $\mathbb{P}^2$ space, both points and lines have two degrees of freedom.

Intersection of lines



Given two lines $\mathbf{l}, \mathbf{l}'$ in the space $\mathbb{P}^2$, their equations can be written as follows:

$$
\begin{aligned}
\mathbf{x}^\intercal \mathbf{l} = 0 \\
\mathbf{x}^\intercal \mathbf{l}' = 0
\end{aligned}
\tag{10}
$$

At this point, the intersection $\mathbf{x}$, regardless of the scale value, signifies a single point and can thus be represented as a multiple of the Cross Product of the two lines $\mathbf{l}, \mathbf{l}'$.

$$\mathbf{x} = \mathbf{l} \times \mathbf{l}' \tag{11}$$

For instance, in $\mathbb{P}^2$ space, the line $x=1$ and the line $y=1$ intersect at $(1,1)$. Using the above formula, the line $x=1$ can be expressed as $-x+1=0 \Rightarrow (-1,0,1)^\intercal$ and the line $y=1$ as $-y+1=0 \Rightarrow (0,-1,1)^\intercal$, resulting in

$$\mathbf{x} = \mathbf{l} \times \mathbf{l}' = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{vmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \tag{12}$$

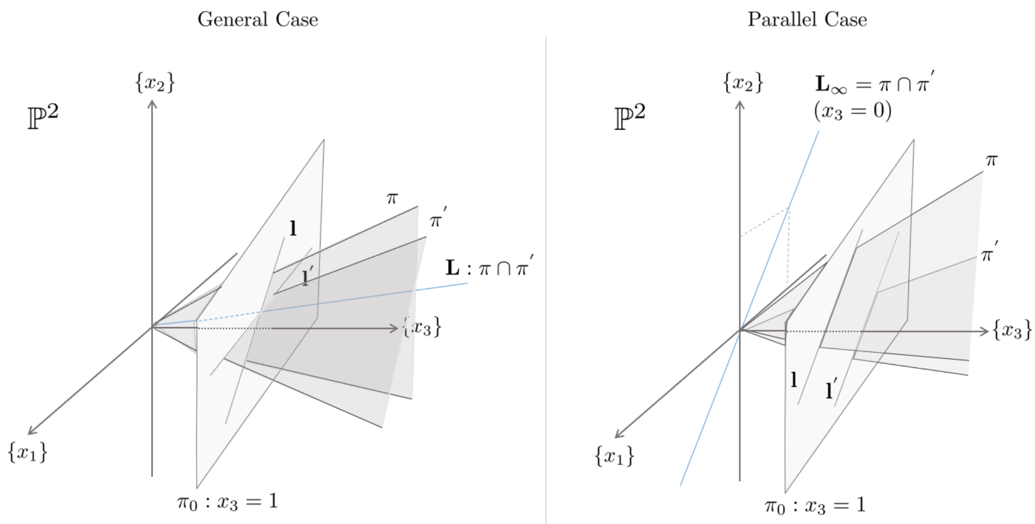
This equation holds true. $(1,1,1)^\intercal$ in $\mathbb{R}^2$ space corresponds to $(1,1)$.

Line joining points

Similarly to the formula for finding the intersection of two lines, in $\mathbb{P}^2$ space, given two points $\mathbf{x}, \mathbf{x}'$, the line $\mathbf{l}$ passing through these two points can be determined as follows:

$$\mathbf{l} = \mathbf{x} \times \mathbf{x}' \tag{13}$$

Ideal points and the line at infinity



Intersection of parallel lines

If two lines $\mathbf{l}, \mathbf{l}'$ are parallel, their intersection point does not meet in $\mathbb{R}^2$ space but does meet in $\mathbb{P}^2$ space.

$$\mathbb{P}^2 = \mathbb{R}^2 \cup \mathbf{l}_\infty \tag{14}$$

The parallel lines $\mathbf{l}, \mathbf{l}'$ can be expressed as follows:

$$\begin{aligned} \mathbf{l} &: (a,b,c)^\intercal \\ \mathbf{l}' &: (a,b,c')^\intercal \end{aligned} \tag{15}$$

The parallel lines intersect at the point $\mathbf{x}_\infty$ located at infinity, therefore

$$\begin{aligned} \mathbf{x}_\infty &= \mathbf{l} \times \mathbf{l}' \\ &= (c'-c) \begin{pmatrix} b \\ -a \\ 0 \end{pmatrix} \sim \begin{pmatrix} b \\ -a \\ 0 \end{pmatrix} \end{aligned} \tag{16}$$

The formula holds. The infinite point $\mathbf{x}_\infty$ transformed into $\mathbb{R}^2$ space becomes $(b/0, -a/0)$, rendering it invalid. Therefore, the point at infinity $\mathbf{x}_\infty = (x,y,0)^\intercal$ in $\mathbb{P}^2$ space does not transform into $\mathbb{R}^2$ space. This indicates that while parallel lines do not meet in Euclidean space, they intersect at infinity in projective space.

For example, in $\mathbb{P}^2$ space, the parallel lines $x=1$ and $x=2$ intersect at infinity. In Homogeneous Notation, this is represented as $-x+1=0 \Rightarrow (-1,0,1)^\intercal$ and $-x+2=0 \Rightarrow (-1,0,2)^\intercal$, resulting in

$$\mathbf{x}_\infty = \mathbf{l} \times \mathbf{l}' = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ -1 & 0 & 1 \\ -1 & 0 & 2 \end{vmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \tag{17}$$

Here, $\mathbf{x}_\infty$ represents a point at infinity in the direction of the y-axis.

Ideal points and the line at infinity

A homogeneous vector $\mathbf{x} = (x_1, x_2, x_3)^\intercal$ corresponds to a point in $\mathbb{R}^2$ space when $x_3 \neq 0$. However, if $x_3 = 0$, the point does not correspond to $\mathbb{R}^2$ space and exists only in $\mathbb{P}^2$ space, where it is known as an Ideal Point or point at infinity. Points at infinity are

$$\mathbf{x}_\infty = \begin{pmatrix} x_1 & x_2 & 0 \end{pmatrix}^\intercal \tag{18}$$

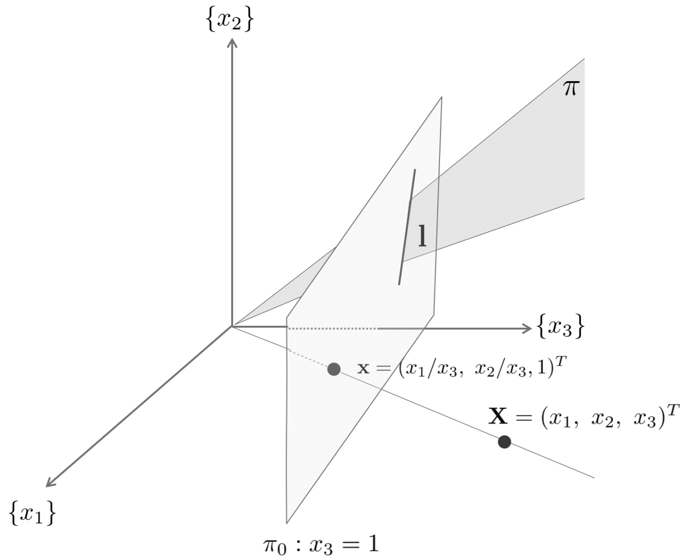
and exist on a specific line known as the line at infinity.

$$\mathbf{l}_\infty = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}^\intercal \tag{19}$$

Thus, $\mathbf{x}_\infty^\intercal \mathbf{l}_\infty = \begin{pmatrix} x_1 & x_2 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}^\intercal = 0$ holds true.

From the previous section, we can see that two parallel lines $\mathbf{l} = (a, b, c)^\intercal$ and $\mathbf{l}'_\infty = (a, b, c')^\intercal$ intersect at the infinite point $\mathbf{x}_\infty = (b, -a, 0)^\intercal$. This reveals that while parallel lines do not intersect in $\mathbb{R}^2$ space, any two distinct lines in $\mathbb{P}^2$ space necessarily intersect at a point.

A model for the projective plane



Geometrically, $\mathbb{P}^2$ represents the set of all lines passing through the origin in the three-dimensional space $\mathbb{R}^3$. Every vector in $\mathbb{P}^2$ can be described as $k(x_1, x_2, x_3)^\intercal$, where the position of a point $(x_1, x_2, x_3)^\intercal$ is determined by the value of k. Since k is a real number, $k = 0$ represents the origin, and $k \neq 0$ represents a set of points forming a line. Conversely, a line passing through the origin in $\mathbb{R}^3$ space corresponds to a point in $\mathbb{P}^2$ space. Extending this, a line $\mathbf{l}$ in $\mathbb{P}^2$ space corresponds to a plane π in $\mathbb{R}^3$ space that includes the origin.

In $\mathbb{P}^2$ space, a point is uniquely defined regardless of the scale value, so the normalized value $(x_1/x_3, x_2/x_3, 1)$ is typically considered the representative value for a point. Therefore, a line in $\mathbb{R}^3$ space passing through the origin and intersecting the plane $x_3 = 1$ corresponds to a point in $\mathbb{P}^2$ space.

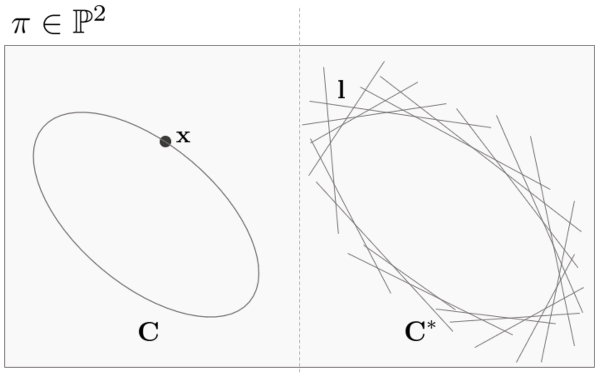
Duality

In $\mathbb{P}^2$ space, points and lines exhibit symmetry (duality). For example, a point $\mathbf{x}$ on a line $\mathbf{l}$ can be expressed in two ways: $\mathbf{x}^\intercal \mathbf{l} = 0$ or $\mathbf{l}^\intercal \mathbf{x} = 0$. Additionally, the point where two lines $\mathbf{l}, \mathbf{l}'$ intersect is

$\mathbf{x} = \mathbf{l} \times \mathbf{l}'$, and the line $\mathbf{l}$ passing through two points $\mathbf{x}, \mathbf{x}'$ can be described as $\mathbf{l} = \mathbf{x} \times \mathbf{x}'$, using the same formula but with the positions of points and lines switched.

Thus, in $\mathbb{P}^2$ space, points and lines have symmetry, meaning that the formula for a line passing through two points is symmetrical to the formula for the intersection point of two lines.

Conics and dual conics



A conic is a curve defined by a quadratic equation in a plane. The general equation is

$$ax^2 + bxy + cy^2 + dx + ey + f = 0 \tag{20}$$

Depending on the coefficients, it can represent various curves such as circles, ellipses, hyperbolas, and parabolas. When represented in Homogeneous Form, it is expressed as

$$ax^2 + bxy + cy^2 + dxz + eyz + fz^2 = 0 \tag{21}$$

When organized in matrix form, it looks like

$$\begin{pmatrix} x & y & z \end{pmatrix} \begin{pmatrix} a & b/2 & d/2 \\ b/2 & c & e/2 \\ d/2 & e/2 & f \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 0 \tag{22}$$

In this form, the symmetric matrix $\begin{pmatrix} a & b/2 & d/2 \\ b/2 & c & e/2 \\ d/2 & e/2 & f \end{pmatrix}$ is referred to as the Conic $\mathbf{C}$.

Five points define a conic

A Conic $\mathbf{C}$ is uniquely determined by five points. The equation for a point on a conic can be rewritten as:

$$\begin{aligned} & ax^2 + bxy + cy^2 + dxz + eyz + fz^2 = 0 \\ & \Rightarrow \begin{pmatrix} x_i^2 & x_i y_i & y_i^2 & x_i & y_i & 1 \end{pmatrix} \mathbf{c} = 0 \end{aligned} \tag{23}$$

Here, $\mathbf{c} = \begin{pmatrix} a & b & c & d & e & f \end{pmatrix} \in \mathbb{R}^6$ represents a vector. Since $\mathbf{c}$ has 5 degrees of freedom,

$$\underbrace{\begin{bmatrix} x_1^2 & x_1 y_1 & y_1^2 & x_1 & y_1 & 1 \\ x_2^2 & x_2 y_2 & y_2^2 & x_2 & y_2 & 1 \\ x_3^2 & x_3 y_3 & y_3^2 & x_3 & y_3 & 1 \\ x_4^2 & x_4 y_4 & y_4^2 & x_4 & y_4 & 1 \\ x_5^2 & x_5 y_5 & y_5^2 & x_5 & y_5 & 1 \end{bmatrix}}_{\mathbf{A}} \mathbf{c} = 0 \tag{24}$$

As stated above, using a total of 5 points, the Null Space vector of matrix $\mathbf{A} \in \mathbb{R}^{5 \times 6}$ becomes the unique solution for $\mathbf{c}$, thus uniquely determining the conic.

Tangent lines to conics

The tangent line $\mathbf{l}$ at a point $\mathbf{x}$ on a Conic $\mathbf{C}$ is expressed as

$$\mathbf{l} = \mathbf{C}\mathbf{x} \tag{25}$$

A conic $\mathbf{C}$ that includes any two lines $\mathbf{l}, \mathbf{m}$ can be written as

$$\mathbf{C} = \mathbf{l}\mathbf{m}^\intercal + \mathbf{m}\mathbf{l}^\intercal \tag{26}$$

Dual conics

The **projective space** $\mathbb{P}^n$ refers to a collection of lines passing through the origin in $\mathbb{R}^{n+1}$ space. In contrast, the **Dual Projective Space** $(\mathbb{P}^n)^\vee$ represents a collection of n-dimensional sub-linear spaces in $\mathbb{R}^n$ space.

An n-dimensional sub-linear space $\mathbf{H}$ is defined as

$$\mathbf{H} = \{\sum_{i=0}^{n} a_i x_i = 0 \mid a_i \neq 0 \text{ for some i}\}. \tag{27}$$

Here $a_0, \cdots, a_n \in \mathbb{P}^n$ can be considered as a single projective space, and a Dual projective space has a symmetric relationship with a projective space.

Given a Conic $\mathbf{C}$ on $\mathbb{P}^2$, the **Dual Conic $\mathbf{C}^*$** on the space $(\mathbb{P}^2)^\vee$ represents a conic and contains information about the tangents to the Conic $\mathbf{C}$. $(\mathbb{P}^2)^\vee$ can parameterize lines on $\mathbb{P}^2$. $\hat{\mathbf{C}}^*$ can be represented as follows.

$$\mathbf{C}^*_{ij} = (-1)^{i+j} \det(\hat{\mathbf{C}}_{ij}) \tag{28}$$

Where $\hat{\mathbf{C}}_{ij}$ denotes the matrix obtained by removing the i-th row and j-th column from $\mathbf{C}_{ij}$.

The necessary and sufficient condition for a line $\mathbf{l}$ to be tangent to a Conic $\mathbf{C}$ is as follows.

$$\mathbf{l}^\intercal \mathbf{C}^* \mathbf{l} = 0 \tag{29}$$

Proof

($\Rightarrow$) Assuming that the Conic $\mathbf{C} \in \mathbb{R}^{3\times 3}$ has rank 3 and is non-singular, it can be expressed as $\mathbf{C}^* = \det(\mathbf{C}^{-1})$. When a point $\mathbf{x} \in \mathbf{C}$ is given on $\mathbf{C}$, the tangent line $\mathbf{l}$ at $\mathbf{x}$ can be represented as $\mathbf{l} = \mathbf{C}\mathbf{x}$. Substituting into the above formula gives

$$\begin{aligned} \mathbf{l}^\intercal \mathbf{C}^* \mathbf{l} &= (\mathbf{C}\mathbf{x})^\intercal \mathbf{C}^* \mathbf{C}\mathbf{x} \\ &= \mathbf{x}^\intercal \mathbf{C}^\intercal \mathbf{C}^* \mathbf{C}\mathbf{x} \\ &= \det(\mathbf{x}^\intercal \mathbf{C}^\intercal \mathbf{x}) \qquad \because \mathbf{C}^* = \det(\mathbf{C}^{-1}) \\ &= 0 \qquad \because \mathbf{x} \in \mathbf{C}, (\mathbf{x}^\intercal \mathbf{C}\mathbf{x})^\intercal = 0 \end{aligned} \tag{30}$$

($\Leftarrow$) When the line $\mathbf{l}$ and Dual Conic $\mathbf{C}^*$ satisfy $\mathbf{l}^\intercal \mathbf{C}^* \mathbf{l} = 0$, it must be proven that $\mathbf{l}$ and $\mathbf{C}^*$ meet at a point $\mathbf{x}$. In this case, the formula $\mathbf{l} = \mathbf{C}\mathbf{x}$ holds.

Since $\mathbf{C}$ is non-singular, it has an inverse, hence $\mathbf{x} = \mathbf{C}^{-1}\mathbf{l}$ can be represented as follows. Therefore, $\mathbf{x}^\intercal \mathbf{l}$ is

$$\begin{aligned} \mathbf{x}^\intercal \mathbf{l} &= (\mathbf{C}^{-1}\mathbf{l})^\intercal \mathbf{l} \\ &= \mathbf{l}^\intercal \mathbf{C}^{-\intercal} \mathbf{l} = 0 \\ &(\mathbf{C}^* \sim \mathbf{C}^{-1} \text{ by assumption.}) \end{aligned} \tag{31}$$

And $\mathbf{x}^\intercal \mathbf{C}\mathbf{x}$ is

$$\begin{aligned} \mathbf{x}^\intercal \mathbf{C}\mathbf{x} &= (\mathbf{C}^{-1}\mathbf{l})^\intercal \mathbf{C}\mathbf{C}^{-1}\mathbf{l} \\ &= \mathbf{l}^\intercal \mathbf{C}^{-\intercal} \mathbf{C}\mathbf{C}^{-1}\mathbf{l} \\ &= \mathbf{l}^\intercal \mathbf{C}^{-1}\mathbf{l} = 0 \\ &(\mathbf{C}^{-\intercal} = \mathbf{C}^{-1} \quad \mathbf{C} is symmetric.) \end{aligned} \tag{32}$$

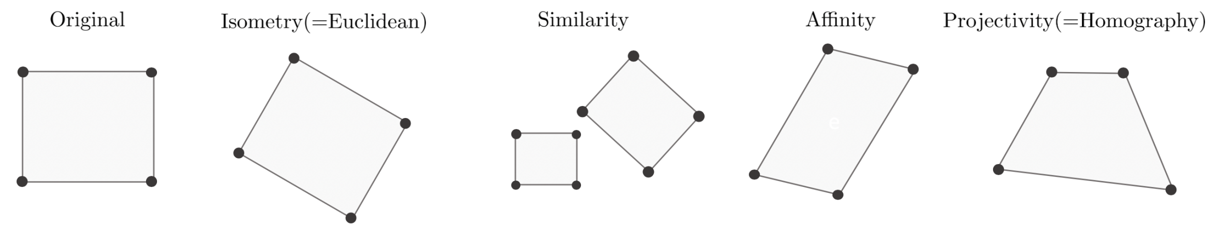
This confirms that $\mathbf{x}$ is the intersection of $\mathbf{l}$ and $\mathbf{C}$.

$$\{\mathbf{x}\} = \mathbf{l} \cap \mathbf{C} \tag{33}$$

Projective transformations

Projective Transformation in $\mathbb{P}^2$ refers to a $\mathbb{P}^2 \Rightarrow \mathbb{P}^2$ mapping defined by a non-singular 3×3 matrix, and it has the property of mapping lines to lines. **Projective Transformation is also known as Collineation, Projectivity, or Homography**.

A hierarchy of transformations



There are various types of transformation matrices depending on what properties are preserved between before and after the transformation.

Class 1: Isometries

If the size of an object remains the same before and after transformation, such transformation is called an Isometry.

$$\mathbf{H}_{iso} = \begin{bmatrix} \mathbf{A} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} \tag{34}$$

Here, $\mathbf{A} \in \mathbb{R}^{2\times 2}$ represents a matrix including 2-dimensional rotation and reflection, and $\mathbf{t} \in \mathbb{R}^2$ is the translation vector for the 2-dimensional object.

Class 2: Similarity transformations

Transformations that include a scale factor s in addition to the Isometry are called Similarity transformations, and **additionally transform the scale along with the translation and rotation of the object**. Here, the $\mathbf{R}$ matrix, where the Reflection property is removed from the existing $\mathbf{A}$ matrix, is used.

$$\mathbf{H}_{S} = \begin{bmatrix} s\mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} \tag{35}$$

Similarity transformation preserves the angle and the ratio of lengths of an object, but not the scale. When two objects are the same up to a Similarity transformation, it means their shapes are identical but differ in scale.

Class 3: Affine transformations

An Affine transformation refers to a transformation matrix without any restrictions on the matrix $\mathbf{A}$ in Isometry transformations. Generally, the object has a different shape after the transformation.

$$\mathbf{H}_{A} = \begin{bmatrix} \mathbf{A} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} \tag{36}$$

$\mathbf{H}_A$ has six degrees of freedom, so it can uniquely determine $\mathbf{H}_A$ from three pairs of corresponding points.

Affine transformations preserve the ratio of lengths of an object and also the property of parallel lines being parallel. Thus, even if the line at infinity $\mathbf{l}_\infty$ is transformed by an Affine transformation, it still remains $\mathbf{l}_\infty$.

$$\mathbf{H}_A(\mathbf{l}_\infty) = \mathbf{l}_\infty \tag{37}$$

Class 4: Projective transformations

Finally, a Projective transformation refers to a transformation matrix where the last row is not necessarily $(0, 0, 1)$. Projective transformations have the property of mapping lines to lines, but do not preserve any of the properties mentioned earlier. Even parallel lines become non-parallel under a Projective transformation, and the ratio of lengths of an object also changes.

$$\mathbf{H}_p = \begin{bmatrix} \mathbf{A} & \mathbf{t} \\ \mathbf{v}^\intercal & v \end{bmatrix} \tag{38}$$

Here, $\mathbf{v} = \begin{bmatrix} v_1 & v_2 \end{bmatrix}^\intercal$ represents an arbitrary 2-dimensional vector, and v also represents an arbitrary scalar value. The Projective transformation matrix $\mathbf{H}_p$ has 8 degrees of freedom, so it can typically be uniquely determined from four pairs of corresponding points.

Decomposition of a projective transformation

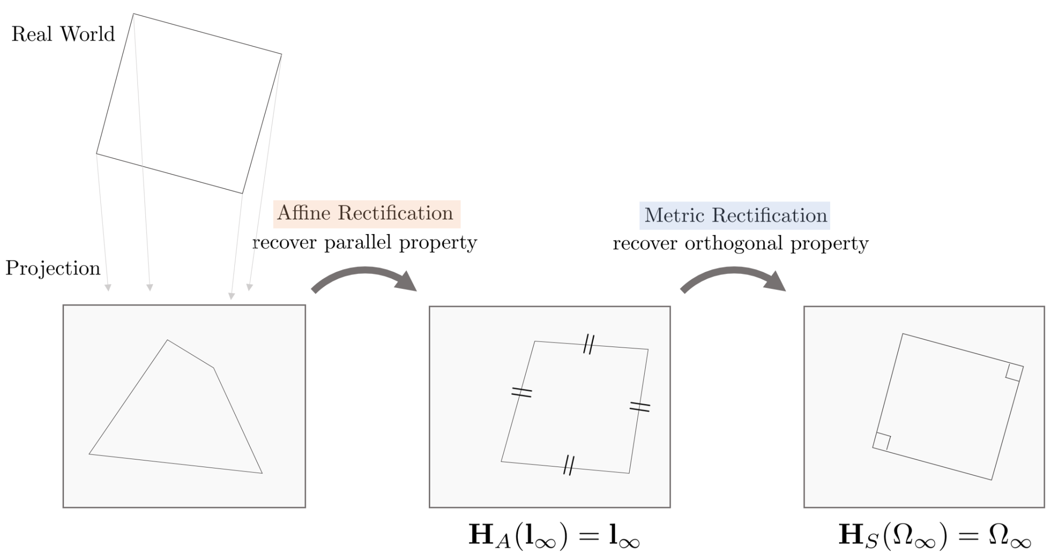
According to the hierarchical structure of the transformation matrices described earlier, a Projective transformation matrix can be expressed as a product of other transformation matrices. Conversely, **a Projective transformation can be decomposed into other transformation matrices.** When an arbitrary Projective transformation $\mathbf{H}_p$ is given,

$$\begin{aligned} \mathbf{H}_p &= \mathbf{H}_S \mathbf{H}_A \mathbf{H}_P \\ &= \begin{bmatrix} s\mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{K} & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{v}^\intercal & v \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{A} & \mathbf{t} \\ \mathbf{v}^\intercal & v \end{bmatrix} \end{aligned} \tag{39}$$

The Projective transformation $\mathbf{H}_p$ can be decomposed into the Similarity transformation $\mathbf{H}_S$, the Affine transformation $\mathbf{H}_A$, and the remaining transformation $\mathbf{H}_P$. Here, $\mathbf{A} = s\mathbf{R}\mathbf{K} + \mathbf{t}\mathbf{v}^\intercal$ and $\mathbf{K}$ represents a normalized upper-triangle matrix with $\det(\mathbf{K}) = 1$. However, such decomposition is only possible when $v \neq 0$ and is uniquely determined when $s > 0$.

$\mathbf{H}^{-1} = \mathbf{H}_P^{-1}\mathbf{H}_A^{-1}\mathbf{H}_S^{-1} = \mathbf{H}'_P\mathbf{H}'_A\mathbf{H}'_S$ also represents the inverse operation of homography in the opposite direction of $\mathbf{H}$. Here, the detailed values of $\mathbf{R}, \mathbf{t}, \mathbf{K}, \mathbf{v}, s, v$ differ between $\mathbf{H}$ and $\mathbf{H}^{-1}$.

Recovery of affine and metric properties from images



When an arbitrary image is given, it is possible to restore the Affine and Metric properties of the image using lines that are parallel and orthogonal in the real world.

The line at infinity

Affine transformation means preserving the Affine property where parallel lines are preserved, and even if the line at infinity $\mathbf{l}_\infty = \begin{bmatrix}0 & 0 & 1\end{bmatrix}^\intercal$ is transformed by an Affine transformation, it still retains the property of being at infinity.

$$\mathbf{H}_A(\mathbf{l}_\infty) = \mathbf{H}_A^{-\intercal}\mathbf{l}_\infty = \begin{bmatrix}\mathbf{A} & \mathbf{t} \\ 0 & 1\end{bmatrix}^{-\intercal}\begin{bmatrix}0 \\ 0 \\ 1\end{bmatrix} = \begin{bmatrix}\mathbf{A}^{-\intercal} & 0 \\ -\mathbf{t}^\intercal\mathbf{A}^{-\intercal} & 1\end{bmatrix}\begin{bmatrix}0 \\ 0 \\ 1\end{bmatrix} = \begin{bmatrix}0 \\ 0 \\ 1\end{bmatrix} = \mathbf{l}'_\infty \tag{40}$$

Thus, $\mathbf{l}_\infty$ remains at infinity even after an Affine transformation.

Recovery of affine properties from images

Recovering affine properties means restoring two lines that are parallel in the real world but appear non-parallel on the image plane due to a projective transformation. If an arbitrary homography $\mathbf{H}$ preserves affine properties, it means that even if the line at infinity $\mathbf{l}_\infty$ is transformed by $\mathbf{H}$, it still remains a line at infinity. That is, if there is a point $\mathbf{x}_\infty$ on the line at infinity, the following holds:

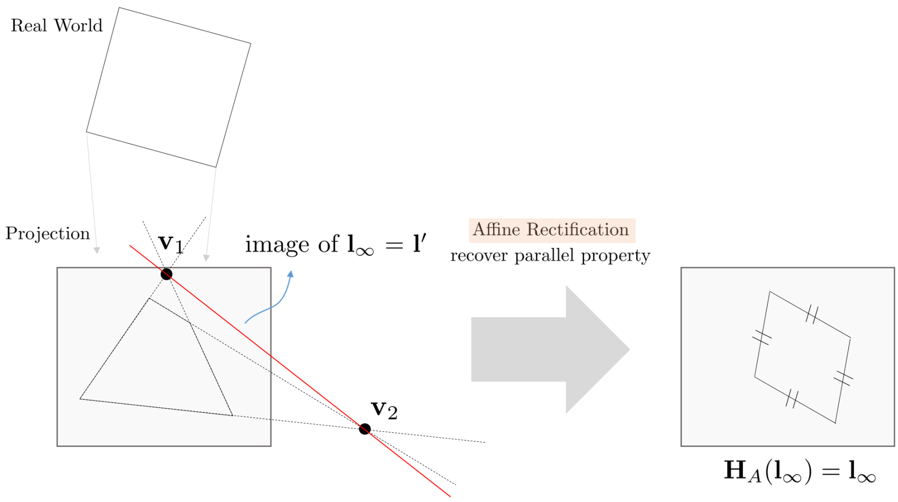
$$\mathbf{H}(\mathbf{x}_\infty) = \mathbf{H}\mathbf{x}_\infty = \mathbf{x}'_\infty \tag{41}$$

A point $\mathbf{x}_\infty$ on the line at infinity is characterized by the last term being 0, such as $\mathbf{x}_\infty = (x, y, 0)^\intercal$, so any homography $\mathbf{H}$ satisfies

$$\mathbf{H}\mathbf{x}_\infty = \begin{bmatrix}\mathbf{A} & \mathbf{t} \\ \mathbf{v} & v\end{bmatrix}\begin{bmatrix}x \\ y \\ 0\end{bmatrix} = \begin{bmatrix}* \\ * \\ 0\end{bmatrix} \tag{42}$$

Therefore, $\mathbf{v} = (0, 0)$ and v becomes a scale constant, allowing transformation to 1.

$$\mathbf{H} = \begin{bmatrix}\mathbf{A} & \mathbf{t} \\ \mathbf{0} & v\end{bmatrix} = \begin{bmatrix}\mathbf{A}/v & \mathbf{t}/v \\ \mathbf{0} & 1\end{bmatrix} \tag{43}$$



However, in images captured through real-world cameras, projective transformations are applied, so the properties of $\mathbf{l}_\infty$ are not preserved and are projected onto the image. Therefore, the process of finding a homography $\mathbf{H}$ that transforms an arbitrary projected line $\mathbf{l}'$ into $\mathbf{l}_\infty$ becomes affine rectification.

$$\mathbf{H}(\mathbf{l}') = \mathbf{H}^{-\intercal}\mathbf{l}' = \mathbf{l}_\infty \tag{44}$$

An arbitrary line can be represented as $\mathbf{l}' = \begin{bmatrix}a & b & c\end{bmatrix}^\intercal$ and $\mathbf{l}_\infty = \begin{bmatrix}0 & 0 & 1\end{bmatrix}^\intercal$, so rephrasing it gives the following.

$$\mathbf{H}(\mathbf{l}') = \mathbf{H}^{-\intercal}\begin{bmatrix}a \\ b \\ c\end{bmatrix} = \begin{bmatrix}0 \\ 0 \\ 1\end{bmatrix} \tag{45}$$

Next, the components of $\mathbf{H}$ must be found. The projective transformation can be divided into three parts as follows:

$$\begin{aligned}\mathbf{H}_p &= \mathbf{H}_S\mathbf{H}_A\mathbf{H}_P \\ &= \begin{bmatrix} s\mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{K} & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{v}^\intercal & v \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{A} & \mathbf{t} \\ \mathbf{v}^\intercal & v \end{bmatrix}\end{aligned} \tag{46}$$

Among these, $\mathbf{H}_P$ does not preserve the property of the line at infinity characteristic of projective transformations. Therefore, the form of $\mathbf{H}$ is as follows.

$$\mathbf{H} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{v}^\intercal & v \end{bmatrix} \tag{47}$$

The $\mathbf{H}$ that transforms $\mathbf{l}^{'}$ to $\mathbf{l}_\infty$ while satisfying the above form is as follows.

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ a & b & c \end{bmatrix} \tag{48}$$

$$\mathbf{H}^{-\intercal}\mathbf{l}^{'} = \mathbf{l}_\infty = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ a & b & c \end{bmatrix}^{-\intercal} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tag{49}$$

The steps for affine rectification are as follows.
1. Obtain the coordinates of two pairs of parallel lines from the real world.
2. Calculate the vanishing point $\mathbf{v}$ for each pair of parallel lines. Since there are two pairs, two $\mathbf{v}_1, \mathbf{v}_2$ are obtained.
3. Obtain the image of the line at infinity $\mathbf{l}^{'} = [a, b, c]^\intercal$ that connects $\mathbf{v}_1$ and $\mathbf{v}_2$.
4. Based on $\mathbf{l}^{'}$, calculate the recovery homography $\mathbf{H}$.

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ a & b & c \end{bmatrix} \tag{50}$$

5. Apply $\mathbf{H}$ to the entire image to complete the affine rectification. The resulting image of affine rectification preserves parallel lines.

Recovery of metric properties from images

The recovery of metric properties means restoring two lines in the real world that are perpendicular but not orthogonal on the image plane due to a projective transformation. This restored image cannot determine the exact scale value (up to similarity, up to scale). Thus, metric rectification means restoring the image to only differ in scale value from the original image. To perform this, it is necessary to use the features of the absolute dual conic $\mathbf{C}^*_\infty$.

- Circular Point

Circular points (or absolute points) $\mathbf{x}_c, \mathbf{x}_{-c}$ are defined as follows.

$$\mathbf{x}_{\pm c} = \begin{bmatrix} 1 \\ \pm i \\ 0 \end{bmatrix} \in \mathbb{CP}^2 \tag{51}$$

- $i = \sqrt{-1}$
- $\mathbb{CP}^2$: complex projective space

If a homography $\mathbf{H}$ preserves the set of circular points, then $\mathbf{H}$ has the property of preserving similarity.

$$\mathbf{H}(\mathbf{x}_{\pm c}) = \mathbf{x}_{\pm c} \quad \text{then, } \mathbf{H} \in \mathbf{H}_s \tag{52}$$

Thus, the form of $\mathbf{H}$ is as follows.

$$\mathbf{H} = \begin{bmatrix} A & t \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} s\mathbf{R} & t \\ 0 & 1 \end{bmatrix} \tag{53}$$

- s : scale factor
- $\mathbf{R}$: rotation matrix

- Dual Conic Properties

When two points $\mathbf{P}$ and $\mathbf{Q}$ exist in $\mathbb{P}^2$ space, the dual conic $\mathbf{C}^*$ tangent to the line connecting the two points can be expressed as follows.

$$\mathbf{C}^* = \mathbf{P}\mathbf{Q}^\intercal + \mathbf{Q}\mathbf{P}^\intercal \tag{54}$$

- $\mathbf{P} = [p_1, p_2, p_3]^\intercal$

This means that $\mathbf{C}^*$ is a dual conic parameterizing the line $\mathbf{l}$ passing through two points $\mathbf{P}$ and $\mathbf{Q}$. The relationship between the dual conic $\mathbf{C}^*$ and the tangent line $\mathbf{l}$ is as follows.

$$\mathbf{l}^\intercal\mathbf{C}^*\mathbf{l} = 0 \tag{55}$$

$$\mathbf{l}^\intercal(\mathbf{P}\mathbf{Q}^\intercal + \mathbf{Q}\mathbf{P}^\intercal)\mathbf{l} = 0 \tag{56}$$

Since line $\mathbf{l}$ includes two points $\mathbf{P}$ and $\mathbf{Q}$, either $\mathbf{P}^\intercal\mathbf{l} = 0$ or $\mathbf{Q}^\intercal\mathbf{l} = 0$ holds, satisfying the above equation.

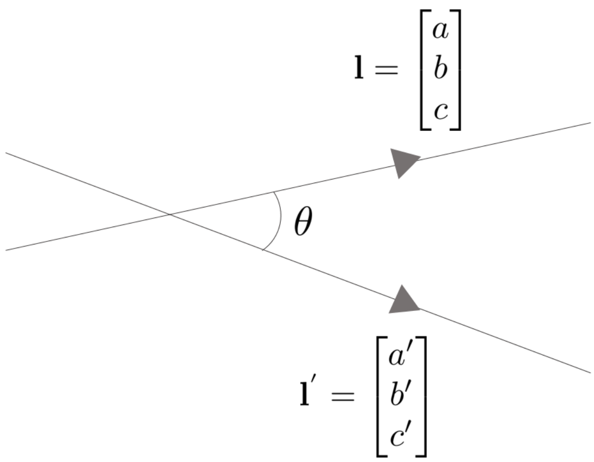
- Absolute Dual Conic

The absolute dual conic $\mathbf{C}^*_\infty$ refers to a dual conic that parameterizes a line passing through two circular points.

$$\mathbf{C}^*_\infty = \mathbf{x}_c\mathbf{x}_{-c}^\intercal + \mathbf{x}_{-c}\mathbf{x}_c^\intercal \tag{57}$$

$$\begin{aligned} \mathbf{C}^*_\infty &= \begin{bmatrix} 1 \\ i \\ 0 \end{bmatrix} \begin{bmatrix} 1 & -i & 0 \end{bmatrix} + \begin{bmatrix} 1 \\ -i \\ 0 \end{bmatrix} \begin{bmatrix} 1 & i & 0 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{aligned} \tag{58}$$

When two lines $\mathbf{l}, \mathbf{l}'$ exist in $\mathbb{P}^2$ space, the angle between the two lines can be expressed as follows.



$$\cos\theta = \frac{aa' + bb'}{\sqrt{a^2 + b^2}\sqrt{a'^2 + b'^2}} \tag{59}$$

When using $\mathbf{C}^*_\infty = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, the equation becomes as follows.

$$\cos\theta = \frac{\mathbf{l}^\intercal\mathbf{C}^*_\infty\mathbf{l}'}{\sqrt{\mathbf{l}^\intercal\mathbf{C}^*_\infty\mathbf{l}}\sqrt{\mathbf{l}'^\intercal\mathbf{C}^*_\infty\mathbf{l}'}} \tag{60}$$

- $aa' + bb' = \mathbf{l}^\intercal \mathbf{C}^*_\infty \mathbf{l}'$
- $\sqrt{a^2+b^2} = \sqrt{\mathbf{l}^\intercal \mathbf{C}^*_\infty \mathbf{l}}$
- $\sqrt{a'^2+b'^2} = \sqrt{\mathbf{l}'^\intercal \mathbf{C}^*_\infty \mathbf{l}'}$

- Homography of Dual Conic

The dual conic and $\mathbf{C}^*$ and the tangent line $\mathbf{l}$ have the following relationship.

$$\mathbf{l}^\intercal \mathbf{C}^* \mathbf{l} = 0 \tag{61}$$

When performing Homography $\mathbf{H} : \mathbb{P}^2 \mapsto \mathbb{P}^2$, the following occurs. Since $\mathbf{H}(\mathbf{l}) = \mathbf{H}^{-\intercal}\mathbf{l}$,

$$(\mathbf{H}^{-\intercal}\mathbf{l})^\intercal \mathbf{H}(\mathbf{C}^*)(\mathbf{H}^{-\intercal}\mathbf{l}) = 0 \tag{62}$$

$$\therefore \mathbf{H}(\mathbf{C}^*) = \mathbf{H}\mathbf{C}^*\mathbf{H}^\intercal \tag{63}$$

The image of the absolute dual conic $\mathbf{w}$ is referred to as $\mathbf{H}(\mathbf{C}^*)$.

- Image of Absolute Dual Conic

If two lines $\mathbf{l}, \mathbf{m}$ in $\mathbb{P}^2$ space are orthogonal, the following equation holds.

$$\mathbf{l}^\intercal \mathbf{w} \mathbf{m} = 0 \tag{64}$$

- $\mathbf{w}$: image of absolute conic $\mathbf{C}^*_\infty$

Since $\mathbf{w} = \mathbf{H}\mathbf{C}^*\mathbf{H}^\intercal$, decomposing an arbitrary projective homography $\mathbf{H}$ results in the following.

$$\begin{aligned} \mathbf{H} &= \mathbf{H}_S\mathbf{H}_A\mathbf{H}_P \\ &= \begin{bmatrix} s\mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{K} & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{v}^\intercal & v \end{bmatrix} \end{aligned} \tag{65}$$

$\mathbf{H}^{-1} = \mathbf{H}_P^{-1}\mathbf{H}_A^{-1}\mathbf{H}_S^{-1} = \mathbf{H}'_P\mathbf{H}'_A\mathbf{H}'_S$ also means the same homography operations. For convenience, $\mathbf{H}'_P\mathbf{H}'_A\mathbf{H}'_S$ is denoted as $\mathbf{H}_P\mathbf{H}_A\mathbf{H}_S$. In this case, the detailed values of $\mathbf{R}, \mathbf{t}, \mathbf{K}, \mathbf{v}, s, v$ of each matrix are different for $\mathbf{H}$ and $\mathbf{H}^{-1}$. Therefore, reversing the decomposition order of $\mathbf{H}$ and multiplying gives the following expansion.

$$\mathbf{H}\mathbf{C}^*\mathbf{H}^\intercal = \mathbf{H}_P\mathbf{H}_A\mathbf{H}_S \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \mathbf{H}_S^\intercal\mathbf{H}_A^\intercal\mathbf{H}_P^\intercal \tag{66}$$

$$\mathbf{H}\mathbf{C}^*\mathbf{H}^\intercal = \mathbf{H}_P\mathbf{H}_A \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \mathbf{H}_A^\intercal\mathbf{H}_P^\intercal \tag{67}$$

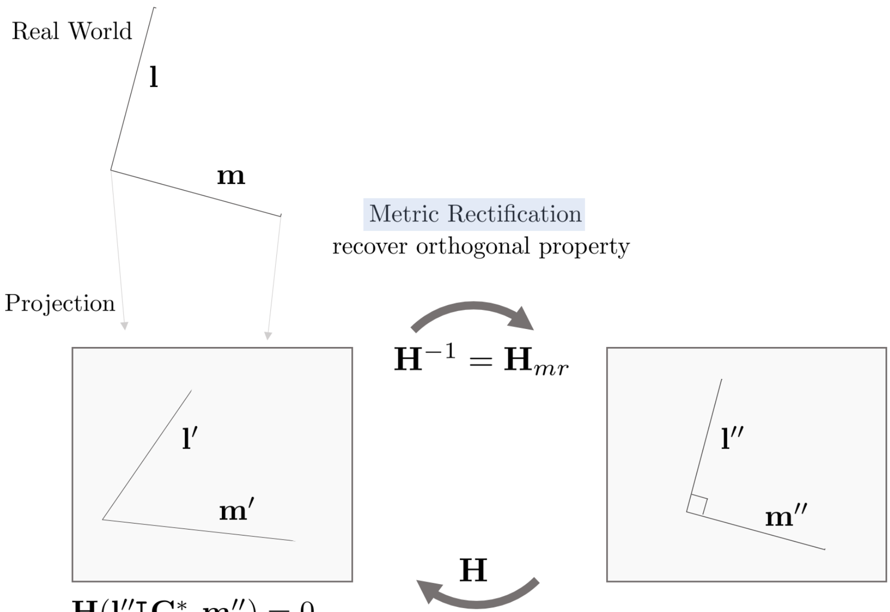
- $\because \mathbf{H}_S \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \mathbf{H}_S^\intercal = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

$$\mathbf{w} = \begin{bmatrix} \mathbf{K}\mathbf{K}^\intercal & \mathbf{K}\mathbf{K}^\intercal\mathbf{v} \\ \mathbf{v}^\intercal\mathbf{K}^\intercal\mathbf{K} & \mathbf{v}^\intercal\mathbf{K}\mathbf{K}^\intercal\mathbf{v} \end{bmatrix} \tag{68}$$

Assuming there is no projective transformation and only a similarity transformation, $\mathbf{v} = 0$ and $\mathbf{w}$ is as follows.

$$\mathbf{w} = \begin{bmatrix} \mathbf{K}\mathbf{K}^\intercal & 0 \\ 0 & 0 \end{bmatrix} \tag{69}$$

- Metric Rectification



$$\mathbf{H}(\mathbf{l}''^{\intercal}\mathbf{C}^*_\infty \mathbf{m}'') = 0$$
$$\mathbf{l}'^{\intercal}\mathbf{H}(\mathbf{C}^*_\infty)\mathbf{m}' = 0$$
$$\therefore \mathbf{l}'^{\intercal}\mathbf{w}\mathbf{m}' = 0$$

As previously mentioned, the image of absolute dual conic by $\mathbf{H}$ can be expressed as $\mathbf{w} = \begin{bmatrix} \mathbf{KK}^\intercal & 0 \\ 0 & 0 \end{bmatrix}$. Therefore, the result of applying the homography transformation $\mathbf{H}$ to $\mathbf{l}''$, $\mathbf{m}''$ in the figure above can be expressed as follows.

$$\mathbf{H}(\mathbf{l}''\mathbf{C}^*_\infty \mathbf{m}'') = \mathbf{l}'\mathbf{w}\mathbf{m}' = 0 \tag{70}$$

- $\mathbf{H}(\mathbf{l}'') = \mathbf{l}'$
- $\mathbf{H}(\mathbf{C}^*_\infty) = \mathbf{w}$
- $\mathbf{H}(\mathbf{m}'') = \mathbf{m}'$

Expanding this further

$$\mathbf{l}' \begin{bmatrix} \mathbf{KK}^\intercal & 0 \\ 0 & 0 \end{bmatrix} \mathbf{m}' = 0 \tag{71}$$

$$\begin{bmatrix} l'_1 & l'_2 \end{bmatrix} \mathbf{KK}^\intercal \begin{bmatrix} m'_1 \\ m'_2 \end{bmatrix} = 0 \tag{72}$$

- $\mathbf{KK}^\intercal \in \mathbb{R}^{2\times 2}$: symmetric matrix & $\det \mathbf{KK}^\intercal = 1$

Therefore, $\mathbf{KK}^\intercal$ can be computed from a pair of perpendicular lines and $\mathbf{w}$ can be determined. When $\mathbf{KK}^\intercal = \mathbf{S}$ is substituted, a symmetric and positive definite matrix can be decomposed as follows.

$$\begin{bmatrix} l'_1 & l'_2 \end{bmatrix} \mathbf{S} \begin{bmatrix} m'_1 \\ m'_2 \end{bmatrix} = 0 \tag{73}$$

$$\mathbf{S} = \mathbf{UDU}^\intercal \tag{74}$$

- $\mathbf{U}$: orthogonal matrix
- $\mathbf{D}$: diagonal matrix

Furthermore, the diagonal matrix $\mathbf{D}$ can be expressed as the product of two matrices $\mathbf{D} = \mathbf{EE}^\intercal$, so it can be rearranged as follows.

$$\mathbf{S} = \mathbf{UE}(\mathbf{UE})^\intercal \tag{75}$$

Next, performing QR decomposition on $\mathbf{UE}$ yields an upper triangle matrix $\mathbf{R}(= \mathbf{K})$ and an orthogonal matrix $\mathbf{Q}$. Expanding this further results in the following.

$$\mathbf{S} = \mathbf{KQQ}^\intercal\mathbf{K}^\intercal = \mathbf{KK}^\intercal \tag{76}$$

- $\mathbf{QQ}^\intercal = \mathbf{I}$

Next, $\mathbf{S}$ is extracted through Cholesky or SVD to find $\mathbf{K}$ and thus determine the final metric rectify homography $\mathbf{H}^{-1} = \mathbf{H}_{mr}$.

$$\mathbf{H} = \begin{bmatrix} \mathbf{K} & 0 \\ 0 & 1 \end{bmatrix} \tag{77}$$

$$\mathbf{H}_{mr} = \mathbf{H}^{-1} = \begin{bmatrix} \mathbf{K} & 0 \\ 0 & 1 \end{bmatrix}^{-1} \tag{78}$$

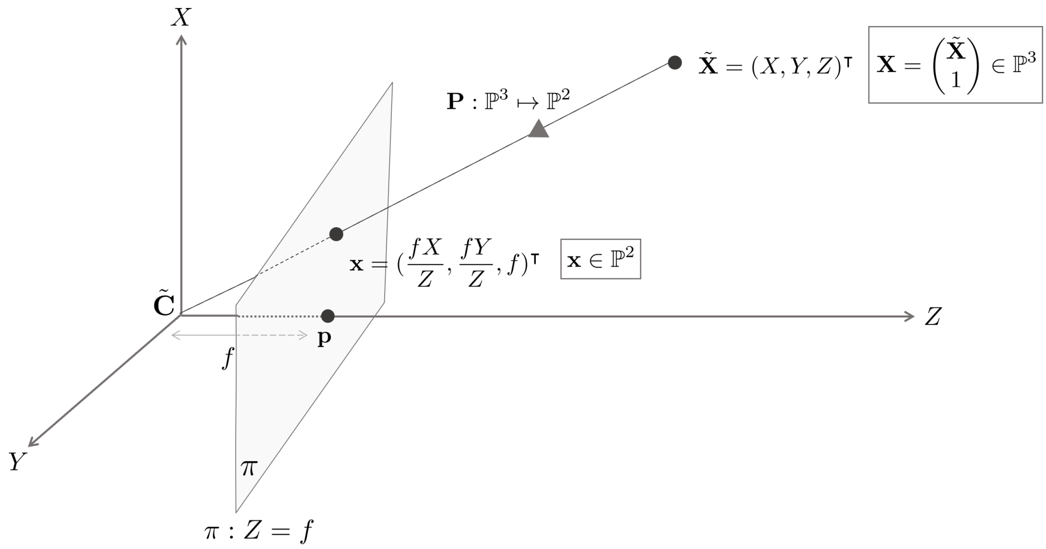
The steps for metric rectification are summarized as follows:

1. Select a pair of perpendicular lines $\mathbf{l}', \mathbf{m}'$ and determine their coordinates.
2. Use the formula $\begin{bmatrix} l'_1 & l'_2 \end{bmatrix} \mathbf{S} \begin{bmatrix} m'_1 \\ m'_2 \end{bmatrix} = 0$ to compute $\mathbf{S} = \mathbf{K}\mathbf{K}^\intercal$.
3. Find $\mathbf{K}$ through Cholesky or SVD, and from this, compute $\mathbf{H}_{mr} = \begin{bmatrix} \mathbf{K} & 0 \\ 0 & 1 \end{bmatrix}^{-1}$.
4. Apply $\mathbf{H}_{mr}$ to the image to perform metric rectification. The restored image will have the same form as the original image except for scale values (up to scale).

3 Camera Models

Finite cameras

The basic pinhole model



A pinhole camera is a mathematical camera modeling method that represents an image by projecting a point $\tilde{\mathbf{X}}$ in $\mathbb{R}^3$ space towards a specific center $\tilde{\mathbf{C}}$, and forming an image at a point $\mathbf{x}$ on the image plane $\pi \in \mathbb{R}^2$ that intersects in the middle. $\tilde{\mathbf{X}}, \tilde{\mathbf{C}}$ are inhomogeneous coordinates representing $\mathbf{X}$.

$$\begin{aligned} \mathbf{X} &= \begin{bmatrix} X & Y & Z & 1 \end{bmatrix}^\intercal \\ \tilde{\mathbf{X}} &= \begin{bmatrix} X & Y & Z \end{bmatrix}^\intercal \\ \mathbf{C} &= \begin{bmatrix} c_x & c_y & c_z & 1 \end{bmatrix}^\intercal \\ \tilde{\mathbf{C}} &= \begin{bmatrix} c_x & c_y & c_z \end{bmatrix}^\intercal \end{aligned} \tag{79}$$

If we consider any $\mathbb{R}^3$ space as a camera coordinate system, the origin of the coordinate system becomes the camera center $\tilde{\mathbf{C}}$. Typically, the image plane π is positioned perpendicular to the Z axis, and the point where the Principal Axis meets the image plane is called the Principal Point $\mathbf{p}$.

Given a point $\tilde{\mathbf{X}} = \begin{bmatrix} X & Y & Z \end{bmatrix}^\intercal$ in 3D space, when only considering the XZ plane, we can calculate the focal length f_x for the X axis, the distance between the image plane π and the camera center $\tilde{\mathbf{C}}$. f_x, f_y imply the pixel aspect ratio, but most sensors produced recently have a 1:1 ratio, so we assume $f = f_x = f_y$.

$$f\frac{Y}{Z} = y \tag{80}$$

Similarly, f can be calculated when looking at the image plane from the YZ plane.

$$f\frac{X}{Z} = x \tag{81}$$

Thus, **the pinhole camera matrix P** can be seen as a linear mapping that projects a point $\tilde{\mathbf{X}} = \begin{pmatrix} X & Y & Z \end{pmatrix}^\intercal \in \mathbb{R}^3$ to a 2D image plane $\pi \in \mathbb{R}^2$.

$$\mathbf{P} : (X,\ Y,\ Z)^\intercal \ \mapsto \ (f\frac{X}{Z},\ f\frac{Y}{Z})^\intercal \tag{82}$$

Central projection using homogeneous coordinates

The pinhole camera matrix $\mathbf{P}$ can also be considered as moving a Homogeneous Point. In other words, the pinhole camera matrix $\mathbf{P}$ projects the point $\mathbf{X} = \begin{pmatrix} X & Y & Z & 1 \end{pmatrix}^\intercal$ in $\mathbb{P}^3$ space to the point $\mathbf{x} = \begin{pmatrix} fX & fY & Z \end{pmatrix}^\intercal$ in $\mathbb{P}^2$ space.

$$\mathbf{P} : \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \mapsto \begin{pmatrix} fX \\ fY \\ Z \end{pmatrix} = \begin{bmatrix} f & & & 0 \\ & f & & 0 \\ & & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \tag{83}$$

This is expressed in matrix form as

$$\mathbf{x} = \mathbf{P}\mathbf{X} \tag{84}$$

At this time, $\mathbf{P} = \mathrm{diag}(f, f, 1)[\mathbf{I} \mid 0]_{3\times 4}$ can be expressed as follows.

Principal point offset

Typically, the Principal Point $\mathbf{p}$ is not the origin of the image plane π. Therefore, to properly correspond the linear mapping through the pinhole camera matrix to the image plane π, it must be corrected as

$$\begin{pmatrix} X & Y & Z \end{pmatrix}^\intercal \mapsto \begin{pmatrix} fX/Z + p_x & fY/Z + p_y \end{pmatrix}^\intercal \tag{85}$$

Expressing this correction in the camera matrix $\mathbf{P}$ at once,

$$\mathbf{P} : \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \mapsto \begin{pmatrix} fX + Zp_x \\ fY + Zp_y \\ 1 \end{pmatrix} = \begin{bmatrix} f & & p_x & 0 \\ & f & p_y & 0 \\ & & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \tag{86}$$

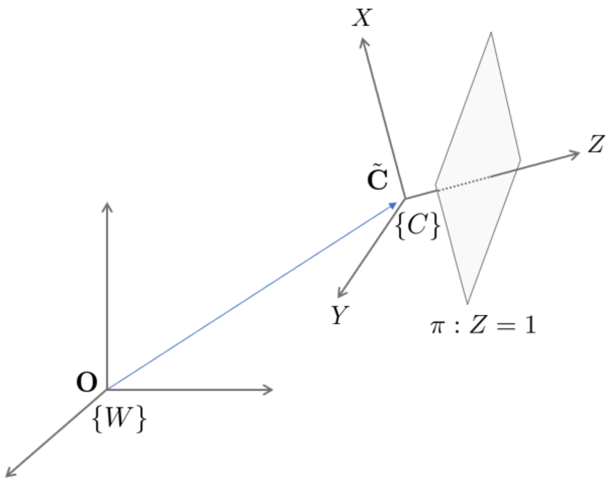
This is represented as $\begin{bmatrix} f & & p_x \\ & f & p_y \\ & & 1 \end{bmatrix}$, which is succinctly denoted as $\mathbf{K}$, and this is called the intrinsic parameter matrix or camera calibration matrix.

$$\mathbf{K} = \begin{bmatrix} f & & p_x \\ & f & p_y \\ & & 1 \end{bmatrix} \tag{87}$$

Consequently, the linear mapping $\mathbf{X} \in \mathbb{P}^3 \mapsto \mathbf{x} \in \mathbb{P}^2$ is possible through the camera matrix $\mathbf{P}$ including the Principal Point Offset.

$$\mathbf{x} = \mathbf{K}[\mathbf{I} \mid 0]\mathbf{X} \tag{88}$$

Camera rotation and translation



- $\tilde{\mathbf{X}}_{\mathbf{C}} = \mathbf{R}(\tilde{\mathbf{X}} - \tilde{\mathbf{C}})$

Typically, the camera coordinate system is not the same as the world coordinate system. Given a world coordinate system $\{W\}$ in $\mathbb{R}^3$ space, the camera coordinate system $\{C\}$, which is rotated by $\mathbf{R}$ and positioned $\mathbf{C} = \begin{pmatrix} c_x & c_y & c_z & 1 \end{pmatrix}^\intercal$ from it, the formula to transform a point $\tilde{\mathbf{X}}$ in the world coordinate system $\{W\}$ to a point $\tilde{\mathbf{X}}_{\mathbf{C}}$ in the camera coordinate system $\{C\}$ is

$$\tilde{\mathbf{X}}_{\mathbf{C}} = \mathbf{R}(\tilde{\mathbf{X}} - \tilde{\mathbf{C}}) \tag{89}$$

The point $\mathbf{X}_{\mathbf{C}}$, represented in Homogeneous Coordinates, when projected onto the image plane π, is expressed as

$$\mathbf{x} = \mathbf{P}\mathbf{X}_{\mathbf{C}} = \mathbf{K}[\mathbf{I} \mid 0]\mathbf{X}_{\mathbf{C}} \tag{90}$$

Detailed expression of $\mathbf{X}_{\mathbf{C}}$ is

$$\begin{aligned} \mathbf{X}_{\mathbf{C}} &= \mathbf{R} \begin{bmatrix} 1 & & & -c_x \\ & 1 & & -c_y \\ & & 1 & -c_z \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{R} & -\mathbf{R}\tilde{\mathbf{C}} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \end{aligned} \tag{91}$$

This can be reformulated and inserted into the formula $\mathbf{x} = \mathbf{K}[\mathbf{I} \mid 0]\mathbf{X}_{\mathbf{C}}$ resulting in

$$\begin{aligned} \mathbf{X}_{\mathbf{C}} &= \mathbf{K}[\mathbf{I} \mid 0]\mathbf{X}_{\mathbf{C}} \\ &= \mathbf{K}[\mathbf{I} \mid 0] \begin{bmatrix} \mathbf{R} & -\mathbf{R}\tilde{\mathbf{C}} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \\ &= \mathbf{K}[\mathbf{R} \mid -\mathbf{R}\tilde{\mathbf{C}}]\mathbf{X} \\ &= \mathbf{K}\mathbf{R}[\mathbf{I} \mid -\tilde{\mathbf{C}}]\mathbf{X} \end{aligned} \tag{92}$$

Typically, $\tilde{\mathbf{X}}_{\mathbf{C}}$ can also often be expressed based on the world coordinate system as $\tilde{\mathbf{X}}_{\mathbf{C}} = \mathbf{R}\tilde{\mathbf{X}} + \mathbf{t}$. In this case, the camera matrix $\mathbf{P}$ can be expressed as

$$\mathbf{P} = \mathbf{K}[\mathbf{R} \mid \mathbf{t}] \tag{93}$$

where the relation $\mathbf{t} = -\mathbf{R}\tilde{\mathbf{C}}$ holds.

CCD cameras

Among the cameras commonly used today, CCD cameras record image coordinates in terms of the number of pixels. Therefore, when the image coordinates are given in mm as $(x, y)[mm]$, in CCD cameras they are represented as $(m_x x,\ m_y y)$. Here m_x, m_y represent the number of pixels in the x-axis or y-axis direction per 1 mm^2. Therefore, to convert a given general camera calibration matrix $\mathbf{K}$ in mm to the coordinate system of a CCD camera, it needs to be transformed as

$$\begin{pmatrix} m_x & & \\ & m_y & \\ & & 1 \end{pmatrix} \mathbf{K} = \begin{pmatrix} m_x & & \\ & m_y & \\ & & 1 \end{pmatrix} \begin{pmatrix} f & & p_x \\ & f & p_y \\ & & 1 \end{pmatrix} = \begin{pmatrix} fm_x & & p_x m_x \\ & fm_y & p_y m_y \\ 0 & 0 & 1 \end{pmatrix} \tag{94}$$

Finite projective camera

When a camera matrix is given as $\mathbf{P} = \mathbf{K}[\mathbf{R} \mid \mathbf{t}]$, it can be represented as

$$\mathbf{P} = \mathbf{KR}[\mathbf{I} \mid -\tilde{\mathbf{C}}] \quad \text{where, } \mathbf{t} = -\mathbf{R}\tilde{\mathbf{C}} \tag{95}$$

This type of camera matrix is referred to as a Finite Projective Camera, and it requires that $\mathbf{KR}$ be a non-singular matrix. Given any non-singular matrix $\mathbf{M} \in \mathbb{R}^{3\times 3}$, it can be decomposed into an upper triangular matrix $\mathbf{K}$ and an orthogonal matrix $\mathbf{R}$ through QR factorization, thus the set of camera matrices are

$$\{\text{set of camera matrix}\} = \{\mathbf{P} = [\mathbf{M} \mid \mathbf{t}] \mid \mathbf{M} \text{ is non-singular } 3\times 3 \text{ matrix.}, \mathbf{t} \in \mathbb{R}^3\} \tag{96}$$

General projective cameras

Unlike Finite Projective Cameras, General Projective Cameras do not require the matrix $\mathbf{M} \in \mathbb{R}^{3\times 3}$ in $\mathbf{P} = [\mathbf{M} \mid \mathbf{t}]$ to be non-singular, and are defined by camera matrices $\mathbf{P}$ with a rank of 3.

The projective camera

Camera anatomy

Camera center

Given any Finite Projective camera matrix $\mathbf{P} = \mathbf{KR}[\mathbf{I} \mid -\tilde{\mathbf{C}}]$

$$\mathbf{PC} = \mathbf{KR}(\mathbf{C} - \mathbf{C}) = \mathbf{0} \tag{97}$$

This means $\mathbf{C} \in \mathbb{R}^4$ represents the camera's center or location in the world coordinate system, and can be determined by finding the Null Space vector of the rank 3 camera matrix $\mathbf{P} \in \mathbb{R}^{3\times 4}$.

Given $\mathbf{PC} = \mathbf{0}$, let's consider a point in the world $\mathbf{X}(\lambda)$ as

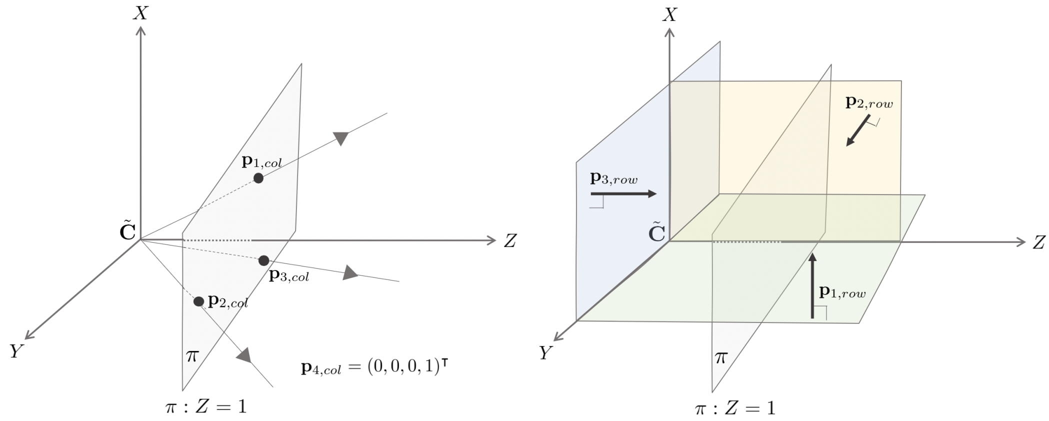
$$\mathbf{X}(\lambda) = \lambda\mathbf{A} + (1-\lambda)\mathbf{C} \tag{98}$$

This represents the line connecting $\mathbf{A}$ and $\mathbf{C}$, and when $\mathbf{X}(\lambda)$ is projected onto the camera, it becomes

$$\mathbf{x} = \mathbf{PX}(\lambda) = \lambda\mathbf{PA} + (1-\lambda)\mathbf{PC} = \lambda\mathbf{PA} \tag{99}$$

This means the line connecting points $\mathbf{A}$ and $\mathbf{C}$ projects onto a single point $\mathbf{x} = \lambda\mathbf{PA}$ in the image plane, regardless of $\mathbf{C}$'s value, illustrating the properties of the camera's center. This property is also true for a General Projective Camera where the Null Space vector of $\mathbf{P}$ represents the camera's center $\mathbf{C}$.

Column vectors



The camera matrix $\mathbf{P}$ can be expressed in terms of column vectors as follows.

$$\begin{aligned} \mathbf{P} &= \begin{bmatrix} \mathbf{p}_{1,col} & \mathbf{p}_{2,col} & \mathbf{p}_{3_col} & \mathbf{p}_{4_col} \end{bmatrix} \\ &\text{where, } \mathbf{p}_{i,col} \in \mathbb{R}^{3\times 1}, \;\; i = 1, \cdots, 4 \end{aligned} \tag{100}$$

Here, $\mathbf{p}_{i,col}, \; i = 1, 2, 3$ represent the vanishing points located at the infinite plane π_∞ for axes X, Y, Z, respectively. And $\mathbf{p}_{4_col} = \mathbf{P}\begin{pmatrix} 0 & 0 & 0 & 1 \end{pmatrix}^\intercal$ represents the origin of the world coordinate system.

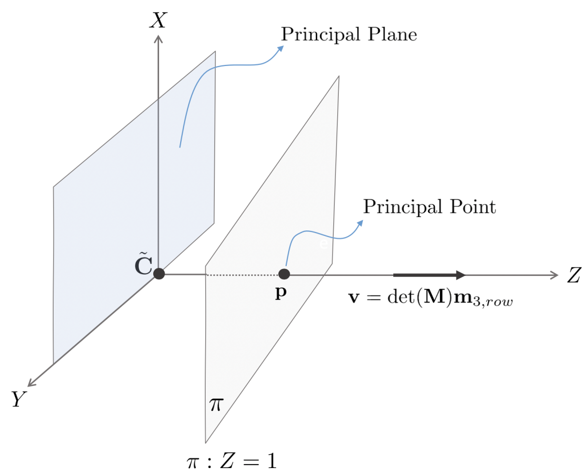
Row vectors

The camera matrix $\mathbf{P}$ can also be expressed using row vectors.

$$\begin{aligned} \mathbf{P} &= \begin{bmatrix} \mathbf{p}_{1,row}^\intercal \\ \mathbf{p}_{2,row}^\intercal \\ \mathbf{p}_{3,row}^\intercal \end{bmatrix} \\ &\text{where, } \mathbf{p}_{i,row} \in \mathbb{R}^{4\times 1}, \;\; i = 1, 2, 3 \end{aligned} \tag{101}$$

Row vector $\mathbf{p}_{i,row}, \; i = 1, 2, 3$ correspond to planes parallel to axes X, Y, Z in the camera coordinate system, respectively.

The principal plane



The principal plane π_{pp} includes the camera's center and is parallel to the image plane. In the camera coordinate system $\{C\}$, it is equivalent to the XY plane with characteristic $Z = 0$. A point $\mathbf{X} \in \pi_{pp}$ on the principal plane meets the image plane π at the line at infinity, thus

$$\mathbf{x} = \mathbf{PX} = \begin{pmatrix} x & y & 0 \end{pmatrix}^\intercal \tag{102}$$

Therefore, the necessary and sufficient condition for a point $\mathbf{X}$ to be located on the principal plane is $\mathbf{p}_{3,row}^\intercal \mathbf{X} = 0$, meaning that the third row vector of the camera matrix $\mathbf{p}_{3,row}$ represents the camera's principal plane.

The Principal Point

The principal point $\mathbf{p}$ refers to the intersection of the principal axis with the image plane π. The principal point $\mathbf{p}$ is located on the image plane π, and the line connecting the camera center $\mathbf{C}$ and the principal point is perpendicular to the image plane.

$$\mathbf{p} - \mathbf{C} \perp \pi \tag{103}$$

The principal point can also be defined as follows. Since the principal plane is the third row vector of the camera matrix $\mathbf{p}_{3,row}$, for a point $\mathbf{X}$ located on the principal plane:

$$\mathbf{p}_{3,row}^\intercal \mathbf{X} = 0 \tag{104}$$

Here, $\mathbf{p}_{3,row} = \begin{pmatrix} \pi_1 & \pi_2 & \pi_3 & \pi_4 \end{pmatrix}^\intercal$ represents the normal vector of the plane $\mathbf{p}_{3,row}$ in Dual Projective Space $(\mathbb{P}^3)^\vee$. The intersection line between the principal plane $\mathbf{p}_{3,row}$ and the plane at infinity π_∞ has the normal vector existing on the plane at infinity as $\begin{pmatrix} \pi_1 & \pi_2 & \pi_3 & 0 \end{pmatrix}^\intercal$. Consequently, the point projected onto the image plane is the principal point $\mathbf{p}$.

$$\mathbf{p} = \mathbf{P} \begin{pmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ 0 \end{pmatrix} \tag{105}$$

The normal vector existing on the plane at infinity $\begin{bmatrix} \pi_1 & \pi_2 & \pi_3 & 0 \end{bmatrix}^\intercal$, passing through the camera center $\mathbf{C}$, is identical to the principal axis.

$$\mathbf{P}\left(\lambda \begin{pmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ 0 \end{pmatrix} + (1-\lambda)\mathbf{C} \right) = \lambda \mathbf{P} \begin{pmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ 0 \end{pmatrix} = \mathbf{p} \tag{106}$$

Thus, projecting the principal axis onto the image plane results in the principal point $\mathbf{p}$. In conclusion, the principal point $\mathbf{p}$ is defined by the first three terms of the third row vector of the camera matrix $\mathbf{P}$, $\mathbf{p}_{3,row} = \begin{bmatrix} \pi_1 & \pi_2 & \pi_3 & p_4 \end{bmatrix}^\intercal$ as $\mathbf{p} = (\pi_1, \pi_2, \pi_3)^\intercal$.

The Principal Axis Vector

When a camera matrix $\mathbf{P} = [\mathbf{M} \mid \mathbf{p}_{4,col}]$ is given, the third row vector of the matrix $\mathbf{M} \in \mathbb{R}^{3\times 3}$, $\mathbf{m}_{3,row}$, represents the principal point. This section considers the direction of the principal axis in the camera coordinate system as equivalent to the $+Z$ direction indicated by $\mathbf{m}_{3,row}$. However, since the camera matrix $\mathbf{P}$ is uniquely determined up to sign, it is impossible to know whether $\mathbf{m}_{3,row}$ or $-\mathbf{m}_{3,row}$ signifies $+Z$.

By multiplying the determinant of $\mathbf{M}$ to $\mathbf{m}_{3,row} = \det(\mathbf{M})\mathbf{m}_{3,row} = (0,0,1)^\intercal$, it always indicates a positive direction. Even if the scale changes from $\mathbf{P} \to k\mathbf{P}$, $\mathbf{v} \to k^4\mathbf{v}$ remains in the same direction. When a typical camera matrix $k\mathbf{P} = k\mathbf{KR}[\mathbf{I} \mid -\tilde{\mathbf{C}}]$ is provided, $\mathbf{M} = k\mathbf{KR}$ and since $\det(\mathbf{R}) > 0$, the direction vector of the axis $\mathbf{v} = \det(\mathbf{M})\mathbf{m}_{3,row}$ remains consistent. Accordingly,

$$\mathbf{v} = \det(\mathbf{M})\mathbf{m}_{3,row} \tag{107}$$

This vector represents the direction vector of the principal axis.

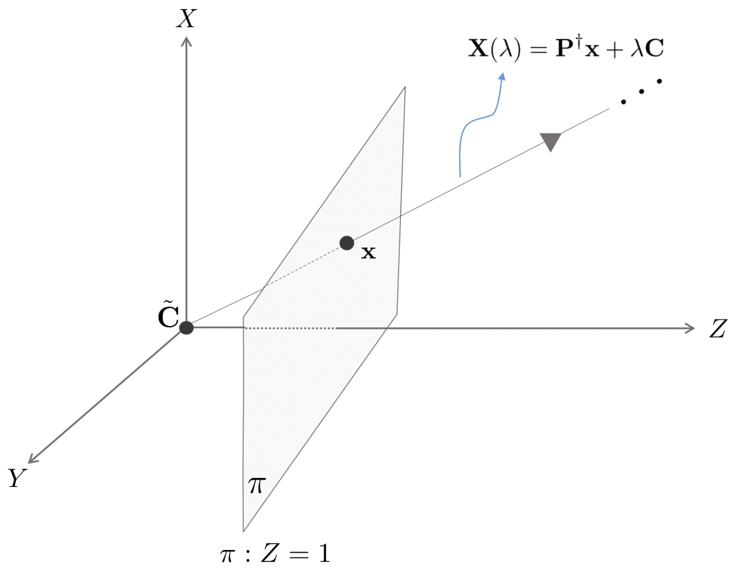
Action of a Projective Camera on Points

Forward Projection

Forward projection, commonly referred to as projection, signifies the operation of transforming a given point $\mathbf{X}$ in the world into a point $\mathbf{x}$ on the image plane. The following formula holds for any camera matrix $\mathbf{P}$:

$$\mathbf{x} = \mathbf{P}\mathbf{X} \tag{108}$$

Back-projection of Points to Rays



Back-projection is the opposite of forward projection and means the operation of transforming a given point $\mathbf{x}$ on the image plane into a line in the world. Typically, the depth value of $\mathbf{x}$ is unknown, hence it cannot be directly transformed into a point $\mathbf{X}$ in the world. Since the rank of any camera matrix $\mathbf{P}$ is 3, a Right Pseudo Inverse $\mathbf{P}^{\dagger}$ exists.

$$\mathbf{P}^{\dagger} = \mathbf{P}^{\intercal}(\mathbf{P}\mathbf{P}^{\intercal})^{-1} \tag{109}$$

Here, $\mathbf{P}\mathbf{P}^{\dagger} = \mathbf{P}\mathbf{P}^{\intercal}(\mathbf{P}\mathbf{P}^{\intercal})^{-1} = \mathbf{I}$ holds true. The line back-projected by $\mathbf{P}^{\dagger}\mathbf{x}$ passes through the camera's center $\mathbf{C}$, thus

$$\mathbf{X}(\lambda) = \mathbf{P}^{\dagger}\mathbf{x} + \lambda\mathbf{C} \tag{110}$$

is represented. Re-projecting the back-projection line results in $\mathbf{P}\mathbf{X}(\lambda) = \mathbf{P}\mathbf{P}^{\dagger}\mathbf{x} + \lambda\mathbf{P}\mathbf{C} = \mathbf{x}$.

In the case of a Finite Projective camera, back-projection can be expressed differently. When an arbitrary Finite Projective camera matrix $\mathbf{P} = [\mathbf{M} \mid \mathbf{p}_{4,col}]$ is given, the camera's center point is represented as $\tilde{\mathbf{C}} = -\mathbf{M}^{-1}\mathbf{p}_{4,col}$. Here, the line back-projected from a point $\mathbf{x}$ on the image plane is tangent to the plane at infinity π_{∞} and $\mathbf{D} = ((\mathbf{M}^{-1}\mathbf{x})^{\intercal}, 0)$, so the back-projection line is

$$\mathbf{X}(\mu) = \mu\begin{pmatrix}\mathbf{M}^{-1}\mathbf{x} \\ 0\end{pmatrix} + \begin{pmatrix}-\mathbf{M}^{-1}\mathbf{p}_{4_col} \\ 1\end{pmatrix} = \begin{pmatrix}\mathbf{M}^{-1}(\mu\mathbf{x} - \mathbf{p}_{4_col}) \\ 1\end{pmatrix} \tag{111}$$

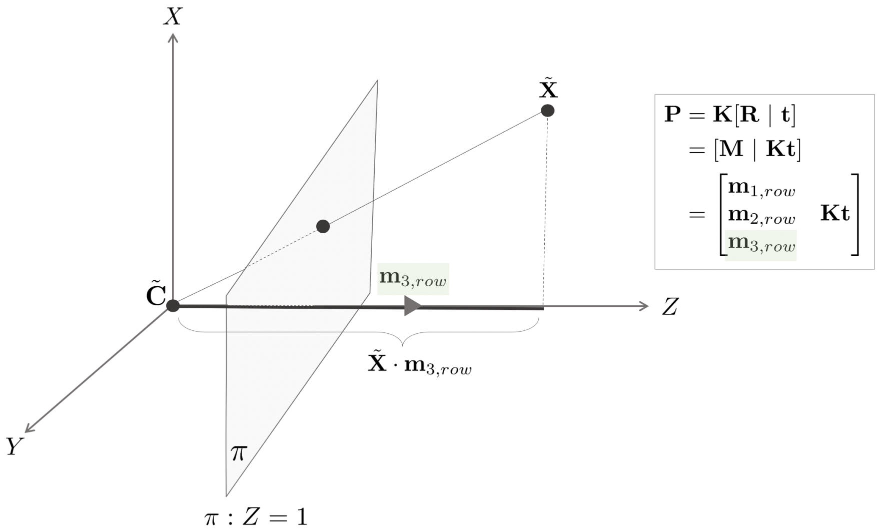
Depth of points

General Projective 카메라 $\mathbf{P}$와 월드 상의 한 점 $\mathbf{X} = \begin{pmatrix}X & Y & Z & 1\end{pmatrix}^{\intercal}$가 주어졌을 때 이를 이미지 평면으로 프로젝션시키면

$$\mathbf{x} = \mathbf{P}\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = \begin{pmatrix} x \\ y \\ w \end{pmatrix} \tag{112}$$

와 같이 하나의 점 $\mathbf{x}$를 얻을 수 있다.

Result 6.1



The depth of a point $\mathbf{X}$ in the world with respect to the camera matrix $\mathbf{P}$ is expressed as

$$\text{depth}(\mathbf{X}; \mathbf{P}) = \frac{\text{sign}(\det(M))w}{\|\mathbf{m}_{3,row}\|} \tag{113}$$

where $\mathbf{m}_{3,row} \in \mathbb{R}^{3\times 3}$ is the third row vector of the matrix $\mathbf{M}$.

Proof

The row vector $\mathbf{m}_{3,row}$ represents the direction of the principal axis, so projecting the world point $\tilde{\mathbf{X}}$ onto $\mathbf{m}_{3,row}$ gives the depth along the Z-axis. The projection onto the principal axis is

$$\text{depth} = \frac{(\tilde{\mathbf{X}} - \tilde{\mathbf{C}})\mathbf{m}_{3.row}}{\|\mathbf{m}_{3,row}\|} \tag{114}$$

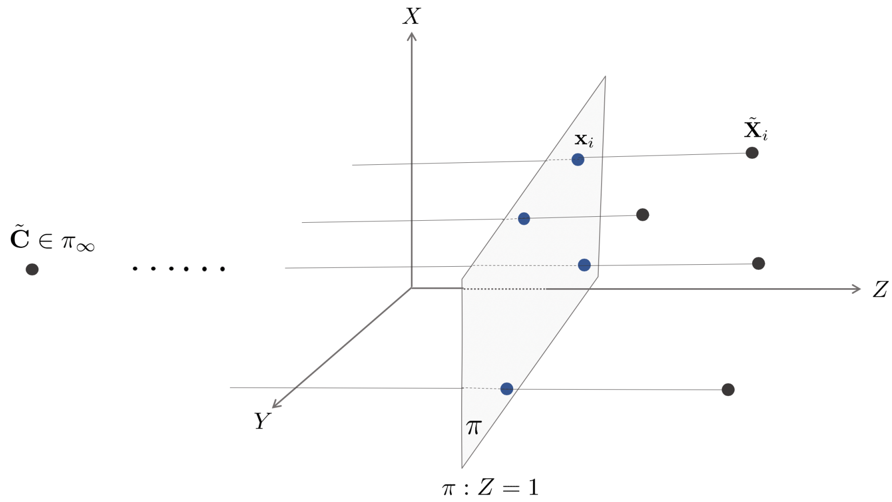
In the case of a Finite Projective camera, $\mathbf{m}_{3,row} = \mathbf{r}_{3,row} = 1$.
The depth value is the third row w of $\mathbf{x} = \mathbf{PX}$, hence

$$\begin{aligned} w &= (\mathbf{PX})_{3,row} \\ &= (\mathbf{P}(\mathbf{X} - \mathbf{C}))_{3,row} \\ &= (\tilde{\mathbf{X}} - \tilde{\mathbf{C}})\mathbf{m}_{3,row} \end{aligned} \tag{115}$$

Thus, considering that the depth value can be located behind the camera depending on the sign of $\det(\mathbf{M})$, it is expressed as follows.

$$\text{depth}(\mathbf{X}; \mathbf{P}) = \frac{\text{sign}(\det(M))w}{\|\mathbf{m}_{3,row}\|} \tag{116}$$

Cameras at infinity

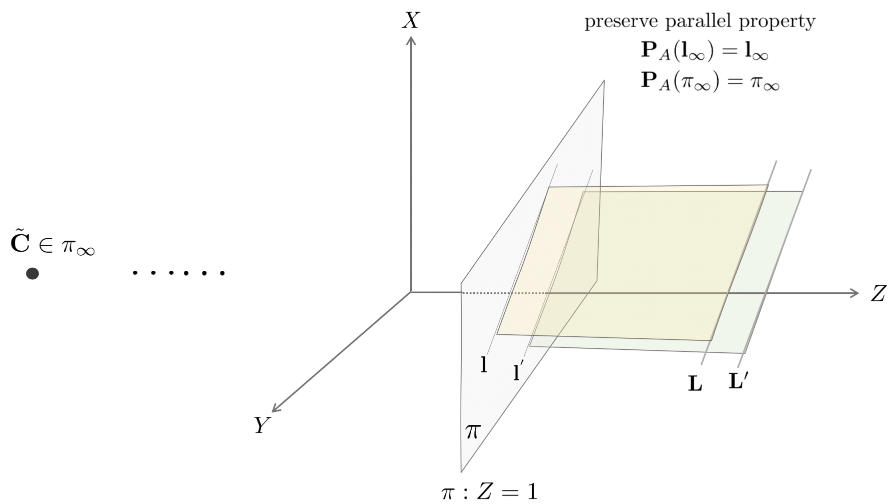


A General Projective camera whose center point C exists on the infinite plane π_∞ is referred to as a camera at infinity.

$$\mathbf{C} = (*,\ *,\ *, 0)^\intercal \in \pi_\infty \tag{117}$$

An equivalent case occurs when the camera matrix $\mathbf{P} = [\mathbf{M} \mid \mathbf{p}_{4,col}]$ is given and matrix $\mathbf{M}$ is singular. Cameras at infinity are broadly classified into Affine cameras and Non-affine cameras.

Definition 6.3

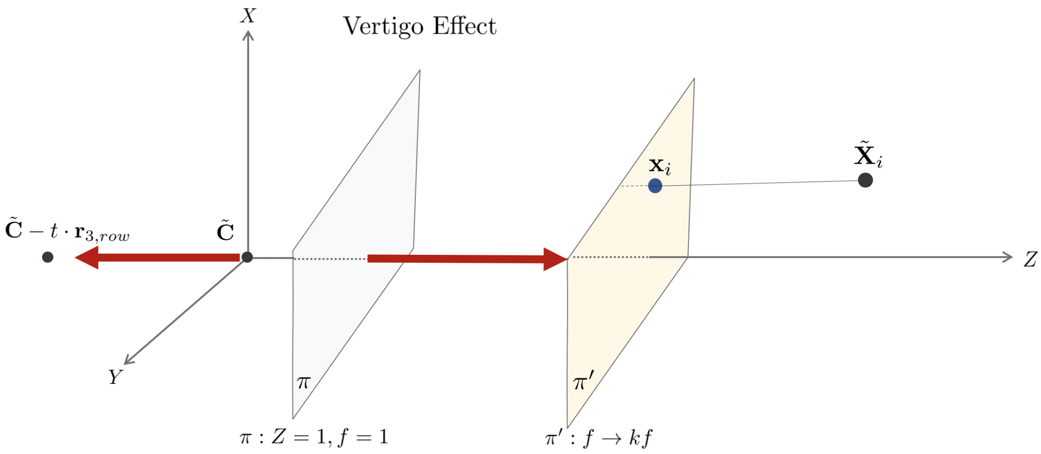


An Affine camera $\mathbf{P}_A$ is defined as a camera that projects the infinite plane to again be the infinite plane.

$$\mathbf{P}_A(\pi_\infty) = \pi_\infty \tag{118}$$

In this case, $\mathbf{P}_A = \begin{bmatrix} * & * & * & * \\ * & * & * & * \\ 0 & 0 & 0 & * \end{bmatrix}$ form.

Affine cameras



Consider a Finite Projective camera matrix $\mathbf{P} = \mathbf{KR}[\mathbf{I} \mid 0]$ with objects existing in the world. If we perform a Zoom In on the object while simultaneously moving the camera in the opposite direction of the principal axis, a Vertigo Effect occurs. The Vertigo Effect is named after the technique first used in the movie Vertigo by director Hitchcock.

To understand this mathematically, reconsider the depth value d of an object in the world given a camera center point $\tilde{\mathbf{C}}$ and a world point $\tilde{\mathbf{X}}$:

$$d = -(\tilde{\mathbf{X}} - \tilde{\mathbf{C}})\mathbf{r}_{3,row} \tag{119}$$

Here, $\mathbf{r}_{3,row}$ is the third row vector of the rotation matrix $\mathbf{R}$ and represents the principal axis. If we denote the distance between the camera center and the world origin as d_0, then for $\tilde{\mathbf{X}} = 0$,

$$d_0 = -\tilde{\mathbf{C}}\mathbf{r}_{3,row} \tag{120}$$

As the camera moves in the opposite direction of the principal axis, the camera center $\tilde{\mathbf{C}}$ becomes

$$\tilde{\mathbf{C}} - t \cdot \mathbf{r}_{3,row} \tag{121}$$

and here, t represents time. As the camera moves backward over time, the camera matrix becomes

$$\begin{aligned} \mathbf{P}_t &= \mathbf{KR}[\mathbf{I} \mid -(\tilde{\mathbf{C}} - t \cdot \mathbf{r}_{3,row})] \\ &= \mathbf{K}\begin{bmatrix} & -\tilde{\mathbf{C}}\mathbf{r}_{1,row} \\ \mathbf{R} & -\tilde{\mathbf{C}}\mathbf{r}_{2,row} \\ & t - \tilde{\mathbf{C}}\mathbf{r}_{3,row} \end{bmatrix} \\ &= \mathbf{K}\begin{bmatrix} & -\tilde{\mathbf{C}}\mathbf{r}_{1,row} \\ \mathbf{R} & -\tilde{\mathbf{C}}\mathbf{r}_{2,row} \\ & d_0 + t \end{bmatrix} \end{aligned} \tag{122}$$

Thus, moving the camera in the opposite direction of the principal axis results in $\mathbf{P}_t$ with only the $(3, 4)$ term being modified to $d_0 + t$. If we define $d_0 + t = d_t$,

$$\mathbf{P}_t = \mathbf{K}\begin{bmatrix} - & - & - & - \\ - & nochange & - & - \\ - & - & - & d_t \end{bmatrix} \tag{123}$$

Next, consider performing a Zoom In. Mathematically, Zoom In involves increasing the focal length f,

$$ZoomIn : f \rightarrow kf \quad \forall k > 0 \tag{124}$$

When expressed as a matrix, Zoom In becomes

$$\mathbf{P} \rightarrow \begin{bmatrix} k & & \\ & k & \\ & & 1 \end{bmatrix} \mathbf{P} \tag{125}$$

When moving the camera in the direction of the principal axis and simultaneously Zooming In by a factor of k, it is possible to create the Vertigo Effect without changing the depth of the object. The appropriate Zoom In factor k is

$$k = d_t/d_0 \tag{126}$$

Ultimately, the camera matrix over time $\mathbf{P}_t$ becomes

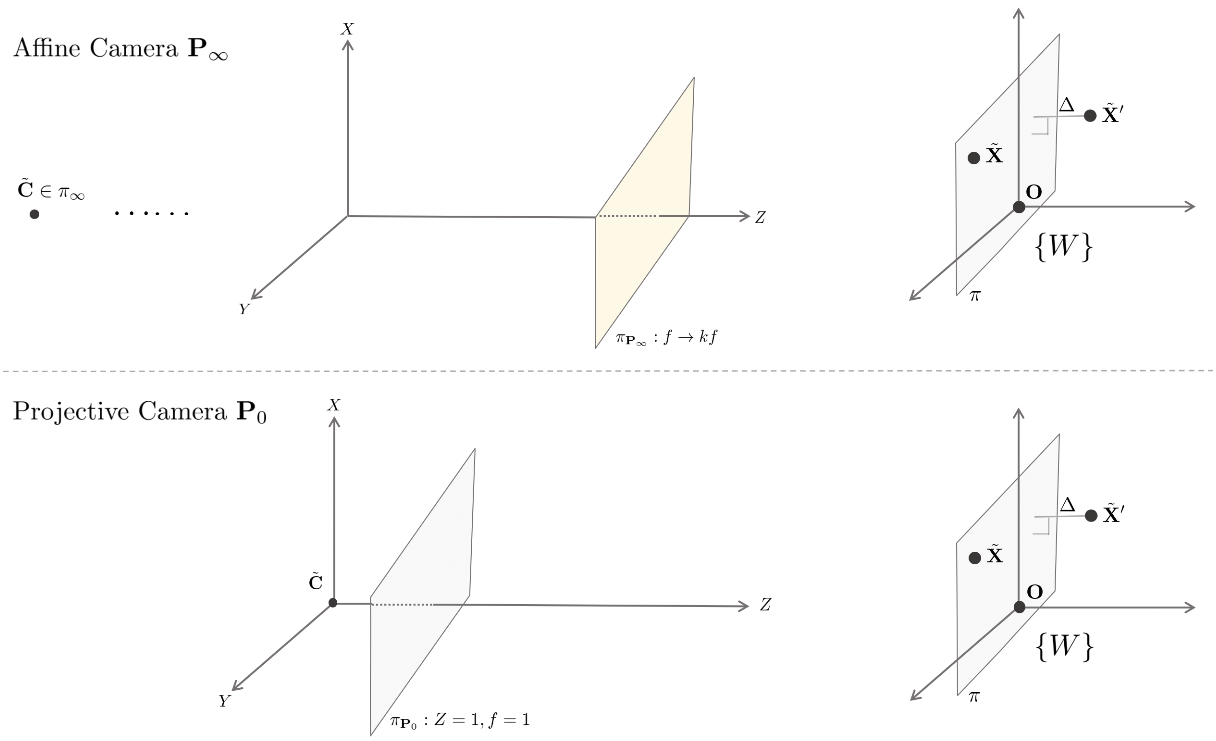
$$\begin{aligned} \begin{bmatrix} d_t/d_0 & & \\ & d_t/d_0 & \\ & & 1 \end{bmatrix} \mathbf{P}_t &= \mathbf{K} \begin{bmatrix} d_t/d_0 & & \\ & d_t/d_0 & \\ & & 1 \end{bmatrix} \left[\begin{array}{c|c} \mathbf{R} & \begin{matrix} * \\ * \\ d_t \end{matrix} \end{array} \right] \\ &= \frac{1}{k}\mathbf{K} \begin{bmatrix} 1 & & \\ & 1 & \\ & & d_0/d_t \end{bmatrix} \left[\begin{array}{c|c} \mathbf{R} & \begin{matrix} * \\ * \\ d_t \end{matrix} \end{array} \right] \\ &= \frac{1}{k}\mathbf{K} \begin{bmatrix} - & - & - & - \\ - & nochange & - & - \\ & d_0/d_t \cdot \mathbf{r}_{3,row} & & d_0 \end{bmatrix} \end{aligned} \tag{127}$$

$\frac{1}{k}$ is a scaling factor and can be omitted. Assuming time progresses infinitely,

$$\mathbf{P}_\infty = \lim_{t\to\infty} \mathbf{P}_t = \mathbf{K} \begin{bmatrix} \mathbf{r}^\intercal_{1,row} & -\mathbf{r}^\intercal_{1,row}\tilde{\mathbf{C}} \\ \mathbf{r}^\intercal_{2,row} & -\mathbf{r}^\intercal_{2,row}\tilde{\mathbf{C}} \\ \mathbf{0}^\intercal & d_0 \end{bmatrix} \tag{128}$$

As the third row of $\mathbf{P}$ consists of $\mathbf{0}^\intercal$, this corresponds to an Affine camera.

Error in employing an affine camera



In this section, the significant differences between capturing the same object with a General Projective camera and an Affine camera are discussed. The General Projective camera is denoted by $\mathbf{P}_0$, the Affine camera by $\mathbf{P}_\infty$, and changes in the camera matrix over time t are denoted by $\mathbf{P}_t$.

When a plane π that includes the origin of the world coordinate system and is perpendicular to the image plane of camera $\mathbf{P}_t$ is given, the points on plane π remain constant in the image obtained by $\mathbf{P}_t$ over time, when performing the Vertigo Effect (zoom in + backward moving).

To prove this, consider a point $\mathbf{X} \in \pi$ on plane π. Since π includes the origin of the world coordinate system, it can be expressed as:

$$\mathbf{X} = \begin{pmatrix} \alpha \mathbf{r}_{1,row} + \beta \mathbf{r}_{2,row} \\ 1 \end{pmatrix} \in \mathbb{R}^4 \tag{129}$$

Projecting this through $\mathbf{P}_t$ yields:

$$\begin{aligned} \mathbf{P}_t\mathbf{X} &= \mathbf{K} \begin{bmatrix} \mathbf{r}_{1,row}^{\intercal} & -\mathbf{r}_{1,row}^{\intercal}\tilde{\mathbf{C}} \\ \mathbf{r}_{2,row}^{\intercal} & -\mathbf{r}_{2,row}^{\intercal}\tilde{\mathbf{C}} \\ d_0/d_t \cdot \mathbf{r}_{3,row} & d_0 \end{bmatrix} \begin{bmatrix} \alpha \mathbf{r}_{1,row} + \beta \mathbf{r}_{2,row} \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} * \\ * \\ d_0 \end{bmatrix} \\ &\because \mathbf{r}_{1,row} \cdot \mathbf{r}_{3,row} = \mathbf{r}_{2,row} \cdot \mathbf{r}_{3,row} = 0 \end{aligned} \tag{130}$$

Therefore, the depth (depth) of the point $\mathbf{X}$ on plane π, which includes the origin of the world coordinate system, is always constant at d_0, thus appearing constant in size over time t. **In other words, $\mathbf{X}$ is transformed into the same point in the image for both the General Projective and Affine cameras.**

$$\mathbf{P}_0\mathbf{X} = \mathbf{P}_t\mathbf{X} = \mathbf{P}_\infty\mathbf{X} \tag{131}$$

However, **if a world point $\mathbf{X}'$, which is Δ away from plane π passing through the origin of the world coordinate system and perpendicular to the image plane, is captured by the two cameras, $\mathbf{P}_0\mathbf{X}' \neq \mathbf{P}_\infty\mathbf{X}'$.** $\mathbf{X}'$ can be expressed as follows.

$$\mathbf{X}' = \begin{pmatrix} \alpha \mathbf{r}_{1,row} + \beta \mathbf{r}_{2,row} + \Delta \mathbf{r}_{3,row} \\ 1 \end{pmatrix} \tag{132}$$

Here, $\mathbf{r}_{3,row}$ represents the camera's principal axis. Projecting $\mathbf{X}'$ with both cameras results in:

$$\mathbf{x}_{\text{proj}} = \mathbf{P}_0\mathbf{X}' = \mathbf{K}\begin{pmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{z}_{\text{proj}} \end{pmatrix} = \mathbf{K}\begin{pmatrix} \alpha - \mathbf{r}_{1,row}^{\intercal}\tilde{\mathbf{C}} \\ \beta - \mathbf{r}_{2,row}^{\intercal}\tilde{\mathbf{C}} \\ d_0 + \Delta \end{pmatrix} \tag{133}$$

$$\mathbf{x}_{\text{affine}} = \mathbf{P}_\infty\mathbf{X}' = \mathbf{K}\begin{pmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{z}_{\text{affine}} \end{pmatrix} = \mathbf{K}\begin{pmatrix} \alpha - \mathbf{r}_{1,row}^{\intercal}\tilde{\mathbf{C}} \\ \beta - \mathbf{r}_{2,row}^{\intercal}\tilde{\mathbf{C}} \\ d_0 \end{pmatrix} \tag{134}$$

$\tilde{z}_{\text{proj}}$ can be calculated as follows.

$$\begin{aligned} \tilde{z}_{\text{proj}} &= [\mathbf{r}_{3,row} | - \mathbf{r}_{3,row}\tilde{\mathbf{C}}]\mathbf{X}' \\ &= [\mathbf{r}_{3,row} | - \mathbf{r}_{3,row}\tilde{\mathbf{C}}] \begin{pmatrix} \alpha \mathbf{r}_{1,row} + \beta \mathbf{r}_{2,row} + \Delta \mathbf{r}_{3,row} \\ 1 \end{pmatrix} \\ &= -\mathbf{r}_{3,row}\tilde{\mathbf{C}} + \Delta \\ &= d_0 + \Delta \end{aligned} \tag{135}$$

The camera calibration matrix $\mathbf{K}$ can be represented as follows.

$$\mathbf{K} = \begin{bmatrix} \mathbf{K}_{2\times 2} & \tilde{\mathbf{x}}_0 \\ \tilde{0}^{\intercal} & 1 \end{bmatrix} \tag{136}$$

Here, $\mathbf{K}_{2\times 2}$ represents an upper-triangular matrix of size 2×2 and $\tilde{\mathbf{x}}_0 = \begin{pmatrix} x_0 & y_0 \end{pmatrix}^{\intercal}$ represents the origin of the image plane. Considering these, the formulas can be reorganized as follows.

$$\begin{aligned}\mathbf{x}_{\text{proj}} &= \begin{pmatrix}\mathbf{K}_{2\times 2}\tilde{\mathbf{x}} + (d_0+\Delta)\tilde{\mathbf{x}}_0 \\ d_0+\Delta\end{pmatrix} \\ \mathbf{x}_{\text{affine}} &= \begin{pmatrix}\mathbf{K}_{2\times 2}\tilde{\mathbf{x}} + d_0\tilde{\mathbf{x}}_0 \\ d_0\end{pmatrix}\end{aligned} \tag{137}$$

Here $\tilde{\mathbf{x}} = \begin{pmatrix}\tilde{x} & \tilde{y}\end{pmatrix}^\intercal$. Calculating the Inhomogeneous coordinates of the points $\mathbf{x}_{\text{proj}}$ and $\mathbf{x}_{\text{affine}}$ results in values obtained by dividing by the last term.

$$\begin{aligned}\tilde{\mathbf{x}}_{\text{proj}} &= \tilde{\mathbf{x}}_0 + \mathbf{K}_{2\times 2}\tilde{\mathbf{x}}/(d_0+\Delta) \\ \tilde{\mathbf{x}}_{\text{affine}} &= \tilde{\mathbf{x}}_0 + \mathbf{K}_{2\times 2}\tilde{\mathbf{x}}/d_0\end{aligned} \tag{138}$$

In conclusion, the difference between the two points projected through the General Projective and Affine cameras is as follows.

$$\tilde{\mathbf{x}}_{\text{affine}} - \tilde{\mathbf{x}}_0 = \frac{d_0+\Delta}{d_0}(\tilde{\mathbf{x}}_{\text{proj}} - \tilde{\mathbf{x}}_0) \tag{139}$$

This equation is called the Discrepancy Equation, and if $\Delta = 0$, i.e., the point is on the plane π which includes the world coordinate system origin and is perpendicular to the image plane, the object captured by the two cameras projects to the same point in the image. This phenomenon can be observed in movies like Vertigo or Jaws, where the protagonist's face remains unchanged while the surroundings zoom out.

4 Computation of the Camera Matrix P

This section describes the method to numerically determine the camera matrix $\mathbf{P}$ using several pairs of corresponding points $(\mathbf{x}_i, \mathbf{X}_i)$ in $\mathbb{P}^3$ space and $\mathbb{P}^2$ space. This method is commonly referred to as Resectioning or Calibration.

Basic equations

When a pair of corresponding points $(\mathbf{x}_i, \mathbf{X}_i)$ is given, the correspondence between the two points is as follows.

$$\mathbf{x}_i = \mathbf{P}\mathbf{X}_i \tag{140}$$

Here, $\mathbf{P}\mathbf{X}_i$ is:

$$\mathbf{P}\mathbf{X}_i = \begin{bmatrix}\mathbf{p}_{1,row}^\intercal \\ \mathbf{p}_{2,row}^\intercal \\ \mathbf{p}_{3,row}^\intercal\end{bmatrix}\mathbf{X}_i \quad = \begin{bmatrix}\mathbf{p}_{1,row}^\intercal\mathbf{X}_i \\ \mathbf{p}_{2,row}^\intercal\mathbf{X}_i \\ \mathbf{p}_{3,row}^\intercal\mathbf{X}_i\end{bmatrix} \tag{141}$$

Using row vectors (row vector), $\mathbf{p}_{i,row} \in \mathbb{R}^{4\times 1}$ is meant. If $\mathbf{x} = \begin{pmatrix}x & y & w\end{pmatrix}^\intercal$, then $\mathbf{x}_i \times \mathbf{P}\mathbf{X}_i = 0$ thus,

$$\mathbf{x}_i \times \mathbf{P}\mathbf{X}_i = \begin{pmatrix} y_i \cdot \mathbf{p}_{2,row}^\intercal\mathbf{X}_i - w_i \cdot \mathbf{p}_{3,row}\mathbf{X}_i \\ w_i \cdot \mathbf{p}_{1,row}^\intercal\mathbf{X}_i - x_i \cdot \mathbf{p}_{3,row}\mathbf{X}_i \\ x_i \cdot \mathbf{p}_{2,row}^\intercal\mathbf{X}_i - y_i \cdot \mathbf{p}_{1,row}\mathbf{X}_i\end{pmatrix} = 0 \tag{142}$$

This can be arranged in the form $\mathbf{A}\mathbf{p} = 0$,

$$\begin{bmatrix}\mathbf{0}^\intercal & -w_i\mathbf{X}_i^\intercal & y_i\mathbf{X}_i^\intercal \\ w_i\mathbf{X}_i^\intercal & \mathbf{0}^\intercal & -x_i\mathbf{X}_i^\intercal \\ -y_i\mathbf{X}_i^\intercal & x_i\mathbf{X}_i^\intercal & \mathbf{0}^\intercal\end{bmatrix}\begin{pmatrix}\mathbf{p}_{1,row} \\ \mathbf{p}_{2,row} \\ \mathbf{p}_{3,row}\end{pmatrix} = 0 \tag{143}$$

Since the last row (row) of the left matrix is linearly dependent, only the first and second rows are represented as

$$\underbrace{\begin{bmatrix}\mathbf{0}^\intercal & -w_i\mathbf{X}_i^\intercal & y_i\mathbf{X}_i^\intercal \\ w_i\mathbf{X}_i^\intercal & \mathbf{0}^\intercal & -x_i\mathbf{X}_i^\intercal\end{bmatrix}}_{\mathbf{A}}\begin{pmatrix}\mathbf{p}_{1,row} \\ \mathbf{p}_{2,row} \\ \mathbf{p}_{3,row}\end{pmatrix} = 0 \tag{144}$$

Here, the matrix $\mathbf{A}$ is of size $\mathbb{R}^{2n\times 12}$, and the vector $\begin{pmatrix} \mathbf{p}_{1,row} \\ \mathbf{p}_{2,row} \\ \mathbf{p}_{3,row} \end{pmatrix}$ is 12×1 in size. Since this equation is in the form $\mathbf{Ap} = 0$, vector $\mathbf{p}$ can be determined using methods such as Singular Value Decomposition (SVD).

Minimal solution

To determine the vector $\mathbf{p} \in \mathbb{R}^{12}$ up to scale, a total of 11 equations are necessary. Using one pair of corresponding points $(\mathbf{x}_i, \mathbf{X}_i)$ generates 2 equations, thus a minimum of 5.5 pairs of corresponding points are required to solve for $\mathbf{p}$. In the case of 5.5 noise-free corresponding point pairs, the rank of matrix $\mathbf{A}$ becomes 11, making the Null Space vector the unique solution vector $\mathbf{p}$.

Over-determined solution

Typically, more than six pairs of corresponding points $(\mathbf{x}_i, \mathbf{X}_i)$ can be obtained, and since the data contains noise, the rank of matrix $\mathbf{A}$ becomes 12. Hence, there is no Null Space, making it impossible to solve for the solution vector $\mathbf{p}$. **In such an Over-determined system of linear equations $\mathbf{Ap} = 0$, the approximate solution vector $\hat{\mathrm{p}}$, which minimizes $\|\mathbf{Ap}\|$ under the condition $\|\mathrm{p}\| = 1$, must be found.**

Degenerate configurations

If more than 5.5 pairs of corresponding points $(\mathrm{x}_i, \mathrm{X}_i)$ are not linearly independent, it becomes impossible to determine the unique solution vector p, and such sets of corresponding points are called Degenerate Configurations. The inability to determine a unique solution vector $\mathbf{p}$ implies that for a point $\mathbf{X}$ in the world,

$$\begin{aligned} &{}^{\exists}\mathbf{P}', \quad \mathbf{P} \neq \mathbf{P}' \\ &\mathbf{PX}_i = \mathbf{P}'\mathbf{X}_i \quad {}^{\forall} i \end{aligned} \tag{145}$$

suggests the existence of another camera matrix $\mathbf{P}'$ that satisfies the conditions, equivalent to saying that $\mathbf{PX} = -\theta\mathbf{P}'\mathbf{X}$ for some constant θ, leading to

$$\underbrace{(\mathbf{P} + \theta\mathbf{P}')}_{\mathbf{P}_\theta}\mathbf{X} = 0 \quad \text{for some } \theta \tag{146}$$

and the set of points $\mathbf{X}$ in the world that satisfy $\mathbf{P}_\theta\mathbf{X} = 0$ cannot differentiate between $\mathbf{P}$ and $\mathbf{P}'$. Such sets are denoted as $\mathcal{S}_\theta$:

$$\mathcal{S}_\theta = \{\mathbf{X} \mid \mathbf{P}_\theta\mathbf{X} = 0\} \tag{147}$$

Points $\mathbf{X}$ that satisfy $\mathcal{S}_\theta$ include:

- All $\mathbf{X}_i$s positioned on a Twisted Cubic
- All $\mathbf{X}_i$s existing on the same plane, including a line through the camera's center point

Twisted Cubic refers to a curve in $\mathbb{P}^3$ space. The set $\mathcal{C}_\theta$ consisting of Twisted Cubic $\mathbf{C}_\theta$:

$$\mathcal{C}_\theta = \{\mathbf{C}_\theta \mid \mathbf{P}_\theta\mathbf{C}_\theta = 0 \ \text{ and } \mathbf{P}_\theta\text{'s rank is 3}\} \tag{148}$$

signifies that $\mathbf{C}_\theta$, typically in the form of a cubic polynomial, lies in the row space of $\mathbf{P}_\theta$. Therefore,

$$\mathbf{C}_\theta = \text{Row } \mathbf{P}_\theta \tag{149}$$

implies that $\mathbf{C}_\theta = \begin{pmatrix} c_1 & c_2 & c_3 & c_4 \end{pmatrix}$, and

$$\det \begin{pmatrix} c_1 & c_2 & c_3 & c_4 \\ - & - & - & - \\ - & \mathbf{P}_\theta & - & - \\ - & - & - & - \end{pmatrix} = 0 \tag{150}$$

results in $\det(2\ 3\ 4)c_1 - \det(1\ 3\ 4)c_2 + \det(1\ 2\ 4)c_3 - \det(1\ 2\ 3)c_4 = 0$. **From this expansion, $\mathbf{C}_\theta$ is expressed as**

$$\mathbf{C}_\theta = (\det(2\ 3\ 4), -\det(1\ 3\ 4), \det(1\ 2\ 4), -\det(1\ 2\ 3)) \tag{151}$$

becoming a Twisted Cubic with each term being a cubic. $\mathbf{P}, \mathbf{P}'$ may share common roots among the terms of $\mathbf{C}_\theta$, and the degrees of each term may fall below cubic. Such cases are referred to as the Degenerate Configuration of $\mathbf{C}_\theta$, and such $\mathbf{C}_\theta$ are not Twisted Cubics.

Line correspondences

When a line $\mathbf{L}$ in the world is projected by the camera matrix $\mathbf{P}$ onto the image plane as line $\mathbf{l}$, unlike points, $\mathbf{l} \neq \mathbf{PL}$.

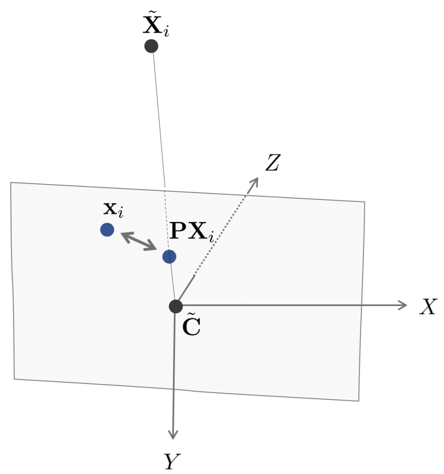
$$\mathbf{x} = \mathbf{PX} \quad \text{but,} \quad \mathbf{l} \neq \mathbf{PL} \tag{152}$$

Since a point $\mathbf{X}$ on the line $\mathbf{L}$ projected by the camera results in point $\mathbf{x}$ existing on line $\mathbf{l}$,

$$\begin{aligned} \mathbf{l}^\intercal \mathbf{x} = \mathbf{l}^\intercal \mathbf{PX} = 0 \\ \Rightarrow \mathbf{Ap} = 0 \end{aligned} \tag{153}$$

A linear equation for vector $\mathbf{p}$ holds as shown above. **Therefore, by using multiple points $\mathbf{X}_i$ on the world line L, a linear equation for vector p holds, and through this, the camera matrix P can be determined.**

Geometric error



As previously explained, an over-determined linear system of the form $\mathbf{Ap} = 0$ can be constructed using pairs of corresponding points $(\mathbf{x}_i, \mathbf{X}_i)$, and an approximate solution vector $\hat{\mathbf{p}}$ can be obtained that minimizes the magnitude of $\|\mathbf{Ap}\|$ while $\|\mathbf{p}\| = 1$. This section explains a method to minimize geometric error in order to obtain a more accurate camera matrix $\mathbf{P}$. **Geometric error refers to the pixel distance between the given $\mathbf{x}_i$ in the image plane and the projected point $\mathbf{PX}_i$ of the world point $\mathbf{X}_i$. In real data, since $\mathbf{x}_i \neq \mathbf{PX}_i$ due to noise, the camera matrix P that minimizes the distance $d(\mathbf{x}_i, \mathbf{PX}_i)$ between the two points must be found.**

$$\min_{\mathbf{P}} d(\mathbf{x}_i, \mathbf{PX}_i)^2 \tag{154}$$

Since $d(\mathbf{x}_i, \mathbf{PX}_i)^2$ is generally non-linear, the optimal camera matrix $\mathbf{P}$ can be determined using non-linear least squares methods such as Gauss-Newton (GN) or Levenberg-Marquardt (LM).

Algorithm 7.1

- **Objective:** Find the MLE (maximum likelihood estimation) value for $\mathbf{P}$ for given pairs of corresponding points $(\mathbf{x}_i, \mathbf{X}_i),\ i = 1, \cdots, 6, \cdots$ to minimize $\sum_i d(\mathbf{x}_i, \mathbf{P}\mathbf{X}_i)^2$.

- **Normalization:** Normalize the image points $\mathbf{x}_i$ through a matrix $\mathbf{T}$ such that $\mathbf{x}_i \rightarrow \bar{\mathbf{x}}_i$ and normalize the world points $\mathbf{X}_i$ through a matrix $\mathbf{U}$ such that $\mathbf{X}_i \rightarrow \bar{\mathbf{X}}_i$. Without performing normalization, direct linear transformation (DLT) results in the last term being very small, equivalent to 1, while other terms are very large, thus not yielding a proper solution.

- **DLT:** Form the normalized pairs of corresponding points into an over-determined system $\bar{\mathbf{A}}\bar{\mathbf{p}} = 0$. Next, find the approximate solution $\bar{\mathbf{p}}$ that minimizes $\|\bar{\mathbf{A}}\bar{\mathbf{p}}\|$ while $\|\bar{\mathbf{p}}\| = 1$ through DLT, and set it as the initial value $\bar{\mathbf{P}}_0$.

- **Minimize geometric error:** Calculate the optimal normalized camera matrix $\bar{\mathbf{P}}$ by minimizing the following geometric error using GN or LM methods.

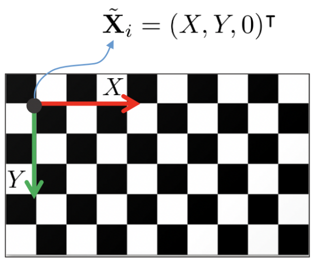
$$\min_{\bar{\mathbf{P}}} \sum_i d(\bar{\mathbf{x}}_i, \bar{\mathbf{P}}\bar{\mathbf{X}}_i)^2 \quad \text{start at } \bar{\mathbf{P}}_0 \tag{155}$$

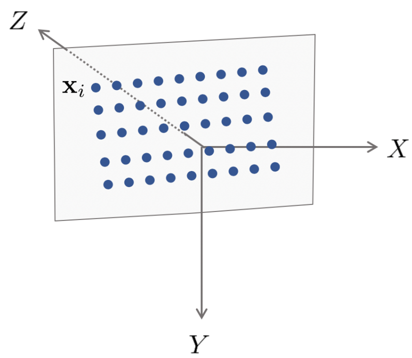
- **Denormalization:** Convert the normalized camera matrix back to the original camera matrix.

$$\mathbf{P} = \mathbf{T}^{-1}\bar{\mathbf{P}}\mathbf{U} \tag{156}$$

This algorithm is commonly referred to as **The Gold Standard algorithm for estimating P.**

Zhang's method





When using the Gold Standard algorithm, instead of using arbitrary pairs of corresponding points, pairs of corresponding points on a checkerboard are used. This algorithm to estimate the camera matrix $\mathbf{P}$ using a checkerboard is called Zhang's Method. When the checkerboard plane π_0 in the world is given, the origin of the world is set to the top left of the checkerboard, and the checkerboard plane is set as the plane $Z = 0$.

$$\pi_0 = \{\mathbf{X} = (X, Y, Z)^\intercal \mid Z = 0\} \tag{157}$$

Accordingly, any point $\mathbf{X}_i$ on the checkerboard plane π_0 becomes a point with $Z = 0$.

$$\mathbf{X}_i = (*,\ *,\ 0)^\intercal \tag{158}$$

Projecting a point $\mathbf{X} = (X,\ Y,\ 0,\ 1)^\intercal$ on the checkerboard results in

$$\begin{aligned}\mathbf{PX} = \mathbf{P}\begin{pmatrix}X\\Y\\0\\1\end{pmatrix} &= \mathbf{K}[\mathbf{R} \mid \mathbf{t}]\begin{pmatrix}X\\Y\\0\\1\end{pmatrix} \\ &= \mathbf{K}[\mathbf{r}_{1,col}\ \mathbf{r}_{2,col}\ \mathbf{t}]\mathbf{X}\end{aligned} \tag{159}$$

Since $z = 0$, the third column vector of the matrix $\mathbf{R}$ becomes 0. The matrix $\mathbf{K}[\mathbf{r}_{1,col}\ \mathbf{r}_{2_col}\ \mathbf{t}] \in \mathbb{R}^{3\times3}$ can be seen as a Homography $\mathbf{H}$ transforming from the checkerboard plane π_0 to the image plane π.

$$\mathbf{H} = \mathbf{K}[\mathbf{r}_{1_col}\ \mathbf{r}_{2_col}\ \mathbf{t}] \tag{160}$$

Since the length and number of patterns on the checkerboard are known, the points $\mathbf{x}_i, i = 1, \cdots$ on the checkerboard plane π_0 can be automatically determined. Next, using a Feature Extraction algorithm, the points $\mathbf{x}'_i, i = 1, \cdots$ seen on the image plane π from π_0 can be determined. Thus, pairs of corresponding points $\mathbf{x}_i \in \pi_0 \leftrightarrow \mathbf{x}'_i \in \pi$ can be obtained. From this, the Homography $\mathbf{H}$ mapping from $\pi_0 \mapsto \pi$ can be calculated.

$$\mathbf{H} = \begin{bmatrix}\mathbf{h}_{1_col} & \mathbf{h}_{2_col} & \mathbf{h}_{3_col}\end{bmatrix} = \mathbf{K}\begin{bmatrix}\mathbf{r}_{1_col} & \mathbf{r}_{2_col} & \mathbf{t}\end{bmatrix} \tag{161}$$

This can be summarized as

$$\mathbf{K}^{-1}\begin{bmatrix}\mathbf{h}_{1_col} & \mathbf{h}_{2_col} & \mathbf{h}_{3_col}\end{bmatrix} = \begin{bmatrix}\mathbf{r}_{1_col} & \mathbf{r}_{2_col} & \mathbf{t}\end{bmatrix} \tag{162}$$

At this time, since the column vectors of the orthogonal matrix $\mathbf{R}$, $\mathbf{r}_{1_col}$ and $\mathbf{r}_{2_col}$, are orthogonal to each other, $\mathbf{r}_{1_col}^\intercal\mathbf{r}_{2_col} = 0$ holds. Using this constraint, $\mathbf{K}^{-1}\mathbf{h}_{1_col} = \mathbf{r}_{1_col}$ and $\mathbf{K}^{-1}\mathbf{h}_{2_col} = \mathbf{r}_{2_col}$, so by the orthogonality condition,

$$\mathbf{h}_{1_col}^\intercal\mathbf{K}^{-\intercal}\mathbf{K}^{-1}\mathbf{h}_{2_col} = 0 \tag{163}$$

It holds. Also, due to the orthogonal matrix condition (up to scale), $\mathbf{r}_{1_col}^\intercal\mathbf{r}_{1_col} = \mathbf{r}_{2_col}^\intercal\mathbf{r}_{2_col}$, so

$$\mathbf{h}_{1_col}^\intercal\mathbf{K}^{-\intercal}\mathbf{K}^{-1}\mathbf{h}_{1_col} = \mathbf{h}_{2_col}^\intercal\mathbf{K}^{-\intercal}\mathbf{K}^{-1}\mathbf{h}_{2_col} \tag{164}$$

The formula holds. Two equations like the above can be obtained from a single checkerboard photograph. The general camera calibration matrix $\mathbf{K}$ is

$$\mathbf{K} = \begin{bmatrix}f_x & s & x_0\\ & f_y & y_0\\ & & 1\end{bmatrix} \tag{165}$$

Therefore, $\mathbf{K}^{-\intercal}\mathbf{K}^{-1}$ also becomes a matrix with 5 variables. Therefore, when at least three Homography $\mathbf{H}_j,\ j = 1, 2, 3$ are given, $\mathbf{K}^{-\intercal}\mathbf{K}^{-1}$ can be determined.

- To determine the matrix $\mathbf{K}$ which has 5 parameters, at least three checkerboard images are acquired. Two equations per image can be obtained, so more than three must be acquired. Homography $\mathbf{H}_j,\ j = 1, 2, 3$ can be obtained for each image, and equations (163), (164) are formulated.
- Set $\mathbf{S} = \mathbf{K}^{-\intercal}\mathbf{K}^{-1}$. Perform Cholesky Decomposition or Singular Value Decomposition (SVD) on matrix $\mathbf{S}$ to find $\mathbf{K}^{-1}$. Matrix $\mathbf{S}$ is a symmetric and Positive Definite matrix, so it can be decomposed as $\mathbf{U}^\intercal\mathbf{D}\mathbf{U}$, and a square root matrix of the diagonal matrix $\mathbf{D} exists$.

$$\text{SVD}(\mathbf{S}) = \mathbf{U}^\intercal\mathbf{D}\mathbf{U} = (\mathbf{U}\sqrt{\mathbf{D}})(\mathbf{U}\sqrt{\mathbf{D}})^\intercal \tag{166}$$

This allows $\mathbf{K}$ to be determined.

- Using the formula $\mathbf{K}^{-1}\mathbf{H} = [\mathbf{r}_{1_col}\ \mathbf{r}_{2_col}\ \mathbf{t}]$, $\mathbf{r}_{1_col}, \mathbf{r}_{2_col}, \mathbf{t}$ are determined, and then

$$\mathbf{r}_{3_col} = \mathbf{r}_1 \times \mathbf{r}_2 \tag{167}$$

is determined. Consequently, the camera's rotation $\mathbf{R}$, translation $\mathbf{t}$, and internal parameter matrix $\mathbf{K}$ can be determined through each Homography $\mathbf{H}_j,\ j = 1, 2, 3$.

Radial Distortion

Real camera images contain radial distortion, which differs from the ideal pinhole camera model, making it impossible to obtain accurate $\mathbf{R}, \mathbf{t}, \mathbf{K}$ without proper calibration. By using actual calibration tools, you can obtain not only $\mathbf{R}, \mathbf{t}, \mathbf{K}$ but also Distortion parameters that correct the radial distortion in the image.

With a small focal length f, the Field of View (FOV) widens, and significant radial distortion occurs near the edges of the image. Conversely, a small f results in a narrow FOV and relatively less distortion.

Let a point on the distorted image plane be $(\breve{u}, \breve{v})$ and a point on the undistorted image plane be (u, v), both in $[pixel]$. Also, let a point on the distorted normalized image plane be $(\breve{x}, \breve{y})$ and a point on the undistorted normalized image plane be (x, y), both in $[mm]$. Being normalized implies that the origin u_0, v_0 of the image plane is zero. The relationship between these normalized points is

$$\begin{aligned} \breve{x} &= x + x(k_1 r^2 + k_2 r^4) \\ \breve{y} &= y + y(k_1 r^2 + k_2 r^4) \quad \text{where, } r^2 = x^2 + y^2 \end{aligned} \tag{168}$$

This indicates that distortion increases with distance from the origin. These normalized points can be represented back on the pixel unit image plane as

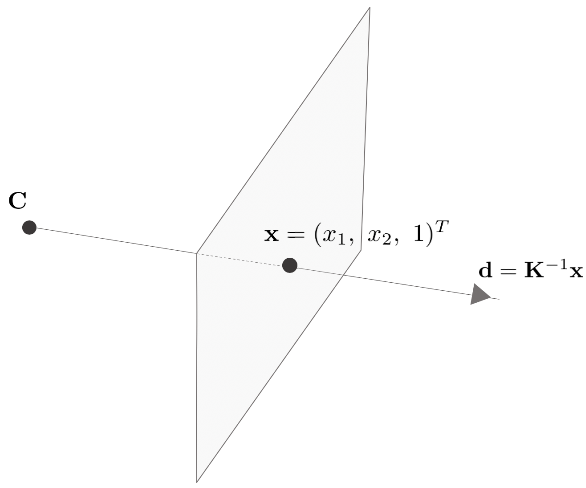
$$\begin{aligned} \breve{u} &= u_0 + \alpha \breve{x} \\ \breve{v} &= v_0 + \beta \breve{v} \end{aligned} \tag{169}$$

Here, u_0, v_0 denote the origin of the image plane and α, β are coefficients that convert points from mm to pixel units. Thus, by determining the radial distortion parameters k_1, k_2, the relationship between the distorted and actual points can be understood.

5 More Single View Geometry

Camera calibration and the image of the absolute conic

Result 8.15

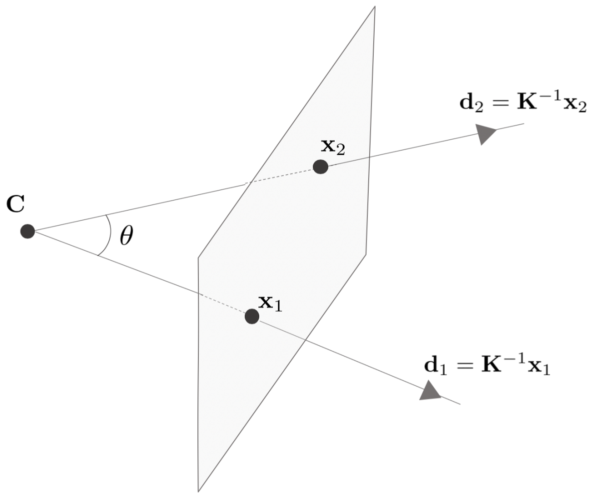


When a camera $\mathbf{C}$ located at the origin back-projects a point $\mathbf{x}$, it creates a line $\mathbf{d}$ passing through the camera center, where $\mathbf{d} = \mathbf{K}^{-1}\mathbf{x}$.

$$\begin{aligned} \mathbf{x} &= \mathbf{P} \begin{bmatrix} \lambda \mathbf{d} \\ 1 \end{bmatrix} \\ &= \mathbf{K}[\mathbf{I}|0] \begin{bmatrix} \lambda \mathbf{d} \\ 1 \end{bmatrix} = \mathbf{K}\mathbf{d} \end{aligned} \tag{170}$$

Therefore, the following formula holds:

$$
\begin{aligned}
\mathbf{x} &= \mathbf{K}\mathbf{d} \\
\mathbf{d} &= \mathbf{K}^{-1}\mathbf{x}
\end{aligned} \tag{171}
$$



The angle between two lines $\mathbf{d}_1, \mathbf{d}_2$ created by back-projecting two points $\mathbf{x}_1, \mathbf{x}_2$ on the image plane can be calculated as follows:

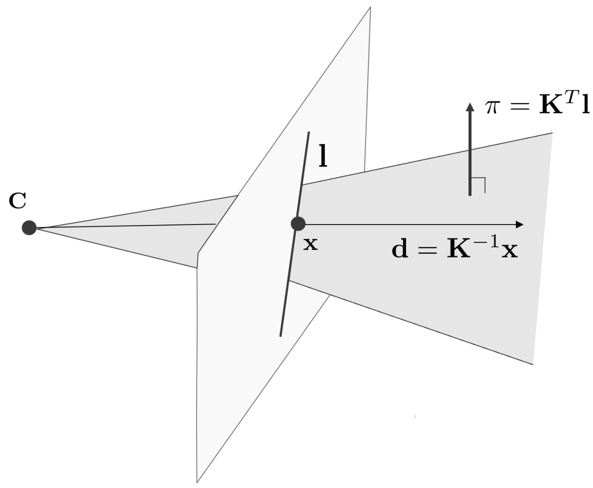
$$
\begin{aligned}
\cos\theta &= \frac{\mathbf{d}_1^\intercal \mathbf{d}_2}{\sqrt{\mathbf{d}_1^\intercal \mathbf{d}_1}\sqrt{\mathbf{d}_2^\intercal \mathbf{d}_2}} = \frac{(\mathbf{K}^{-1}\mathbf{x}_1)^\intercal(\mathbf{K}^{-1}\mathbf{x}_2)}{\sqrt{(\mathbf{K}^{-1}\mathbf{x}_1)^\intercal(\mathbf{K}^{-1}\mathbf{x}_1)}\sqrt{(\mathbf{K}^{-1}\mathbf{x}_2)^\intercal(\mathbf{K}^{-1}\mathbf{x}_2)}} \\
&= \frac{\mathbf{x}_1^\intercal(\mathbf{K}^{-\intercal}\mathbf{K}^{-1})\mathbf{x}_2}{\sqrt{\mathbf{x}_1^\intercal\mathbf{K}^{-\intercal}\mathbf{K}^{-1}\mathbf{x}_1}\sqrt{\mathbf{x}_2^\intercal\mathbf{K}^{-\intercal}\mathbf{K}^{-1}\mathbf{x}_2}}
\end{aligned} \tag{172}
$$

Here, **the image of the Absolute Conic is $\mathbf{K}^{-\intercal}\mathbf{K}^{-1}$.**

Result 8.16

When a point $\mathbf{x}$ lies on a line $\mathbf{l}$ on the image plane, the following formula holds:

$$
\mathbf{x}^\intercal \mathbf{l} = 0 \tag{173}
$$



When $\mathbf{l}$ is back-projected, it creates a plane π and when $\mathbf{x}$ is back-projected, it creates a line $\mathbf{d} = \mathbf{K}^{-1}\mathbf{x}$, and the relationship $(\mathbf{K}^{-1}\mathbf{x})^\intercal\pi = 0$ holds. Rearranging this gives $\mathbf{x}^\intercal(\mathbf{K}^{-\intercal}\pi) = 0$. According to the formula $\mathbf{x}^\intercal\mathbf{l} = 0$, the following formula holds conclusively:

$$
\begin{aligned}
\mathbf{K}^{-\intercal}\pi &= \mathbf{l} \\
\pi &= \mathbf{K}^{\intercal}\mathbf{l}
\end{aligned}
\tag{174}
$$

The image of the absolute conic

When there is an infinite plane π_∞, and an infinite point $\mathbf{X}_\infty = (\mathbf{d}^\intercal, 0)^\intercal$ exists on π_∞, projecting it with camera $\mathbf{P} = \mathbf{KR}[\mathbf{I}| - \tilde{\mathbf{C}}]$ results in the following:

$$
\mathbf{x} = \mathbf{PX}_\infty = \mathbf{KR}[\mathbf{I}| - \tilde{\mathbf{C}}] \begin{bmatrix} \mathbf{d} \\ 0 \end{bmatrix} = \mathbf{KRd} \tag{175}
$$

Thus, a Homography $\mathbf{H}$ exists such that $\mathbf{x} = \mathbf{Hd}$, where Homography $\mathbf{H}$ is $\mathbf{H} = \mathbf{KR}$. The Absolute Conic Ω_∞ at infinity is $\mathbf{I}_3 \in \pi_\infty$. Transforming Ω_∞ by Homography results in

$$
\begin{aligned}
\mathbf{H}(\Omega_\infty) &= \mathbf{H}^{-\intercal}\mathbf{I}_3\mathbf{H}^{-1} \\
&= (\mathbf{KR})^{-\intercal}\mathbf{I}_3(\mathbf{KR})^{-1} \\
&= \mathbf{K}^{-\intercal}\mathbf{R}^{-\intercal}\mathbf{R}^{-1}\mathbf{K}^{-1} \\
&= \mathbf{K}^{-\intercal}\mathbf{K}^{-1}
\end{aligned}
\tag{176}
$$

Result 8.17

Accordingly, **the Image of Absolute Conic $\mathbf{w} = \mathbf{K}^{-\intercal}\mathbf{K}^{-1}$** is established.

When two points $\mathbf{x}_1, \mathbf{x}_2$ on the image plane are back-projected, the angle between the two lines is

$$
\cos\theta = \frac{\mathbf{x}_1^\intercal \mathbf{w} \mathbf{x}_2}{\sqrt{\mathbf{x}_1^\intercal \mathbf{w} \mathbf{x}_1}\sqrt{\mathbf{x}_2^\intercal \mathbf{w} \mathbf{x}_2}} \tag{177}
$$

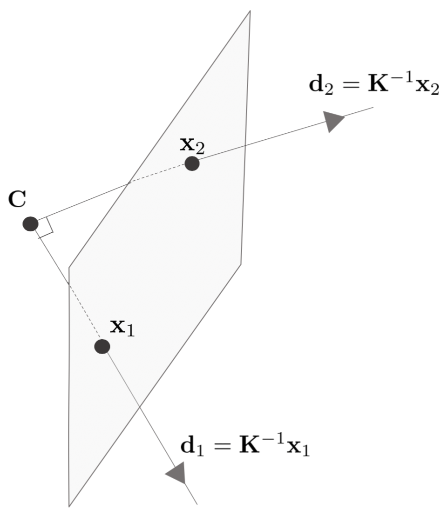
and transforming it by Homography results in

$$
\cos\theta = \frac{(\mathbf{Hx}_1)^\intercal \mathbf{H}^{-\intercal}\mathbf{w}\mathbf{H}^{-1}(\mathbf{Hx}_2)}{\sqrt{*}\sqrt{*}} \tag{178}
$$

Thus, the angles are preserved even after the Homography transformation. If the two lines $\mathbf{K}^{-1}\mathbf{x}_1$ and $\mathbf{K}^{-1}\mathbf{x}_2$ are orthogonal, the following formula holds:

$$
\mathbf{x}_1^\intercal \mathbf{w} \mathbf{x}_2 = 0 \tag{179}
$$

Orthogonality and w



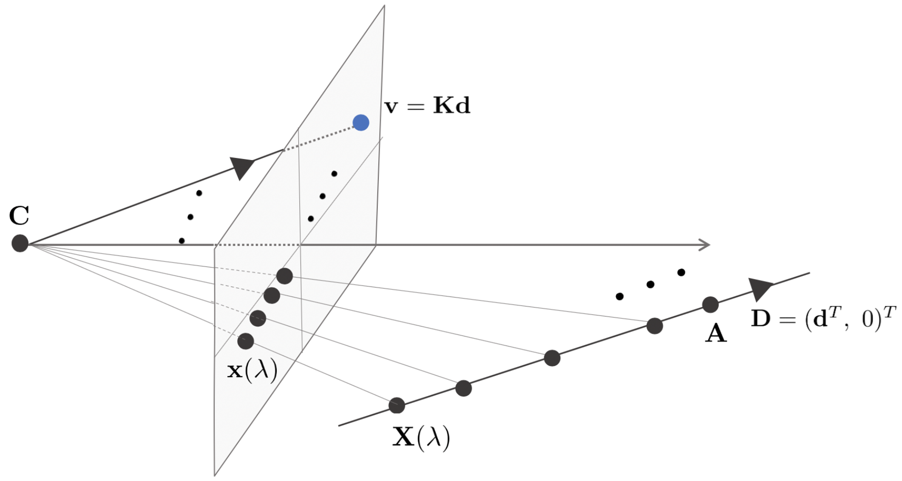
When two points $\mathbf{x}_1, \mathbf{x}_2$ on the image plane are back-projected, the two lines $\mathbf{K}^{-1}\mathbf{x}_1, \mathbf{K}^{-1}\mathbf{x}_2$ are created. If these two lines are orthogonal, the formula $\mathbf{x}_1^\intercal \mathbf{w}\mathbf{x}_2 = 0$ holds. Also, if $\mathbf{x}_1$ is included in line $\mathbf{l}$, then $\mathbf{x}_1^\intercal \mathbf{l} = 0$ holds.

Result 8.19

Combining the two formulas results in

$$\begin{aligned} &\mathbf{x}_1^\intercal \mathbf{w}\mathbf{x}_2 = 0 \\ &\mathbf{x}_1^\intercal \mathbf{l} = 0 \\ &\therefore \mathbf{l} = \mathbf{w}\mathbf{x}_2 \end{aligned} \tag{180}$$

Vanishing points and vanishing lines



When a point $\mathbf{A}$ and the direction of a line $\mathbf{D} = (\mathbf{d}^\intercal, 0)^\intercal$ exist in the world, a point $\mathbf{X}(\lambda)$ on the line is defined as follows:

$$\mathbf{X}(\lambda) = \mathbf{A} + \lambda\mathbf{D} = \begin{bmatrix} \tilde{\mathbf{A}} + \lambda\mathbf{d} \\ 1 \end{bmatrix} \tag{181}$$

The point $\mathbf{x}(\lambda) = \mathbf{P}\mathbf{X}(\lambda)$, where $\mathbf{P} = \mathbf{K}[\mathbf{I}|0]$ projected on the image plane is defined as follows:

$$\mathbf{x}(\lambda) = \mathbf{P}\mathbf{X}(\lambda) = \mathbf{P}\mathbf{A} + \lambda\mathbf{P}\mathbf{D} = \mathbf{a} + \lambda\mathbf{K}\mathbf{d} \tag{182}$$

Here, $\mathbf{a}$ represents the Image of $\mathbf{A}$.

Result 8.20

Conclusively, the vanishing point $\mathbf{v}$ is defined as follows:

$$\begin{aligned} &\mathbf{v} = \lim_{\lambda\to\infty} \mathbf{x}(\lambda) = \lim_{\lambda\to\infty} (\mathbf{a} + \lambda\mathbf{K}\mathbf{d}) = \mathbf{K}\mathbf{d} \\ &\mathbf{v} = \mathbf{K}\mathbf{d} \end{aligned} \tag{183}$$

Camera rotation from vanishing points

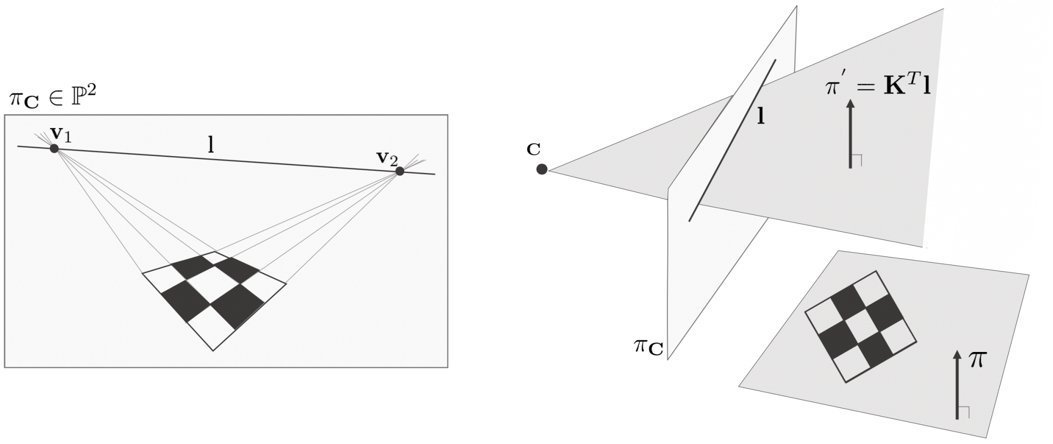
Using vanishing points, the rotation of the camera can be calculated. If there are vanishing points $\mathbf{v}_1$ from Image 1 and $\mathbf{v}_2$ from Image 2, the direction of $\mathbf{v}_1$ is $\mathbf{d}_1 = \mathbf{K}^{-1}\mathbf{v}_1$ and the direction of $\mathbf{v}_2$ is $\mathbf{d}_2 = \mathbf{K}^{-1}\mathbf{v}_2$.

Assuming the values of $\mathbf{K}, \mathbf{v}_1, \mathbf{v}_2$ are all known, the values of $\mathbf{d}_1, \mathbf{d}_2$ can be calculated and the two direction vectors have the following relationship:

$$\mathbf{d}_2 = \mathbf{R}\mathbf{d}_1 \tag{184}$$

In this case, **since the degree of freedom of the rotation matrix R is 3, the rotation matrix can be recovered by using more than two pairs of vanishing points**.

Vanishing Lines



Lines connecting two or more vanishing points $\mathbf{v}_i, i = 1, 2, \cdots$ are called vanishing lines $\mathbf{l}$. For instance, consider an image with a checkerboard leading to two vanishing points $\mathbf{v}_1, \mathbf{v}_2$. In this case, **the line connecting the two vanishing points $\mathbf{v}_1, \mathbf{v}_2$ is referred to as the vanishing line l of the checkerboard plane π.**

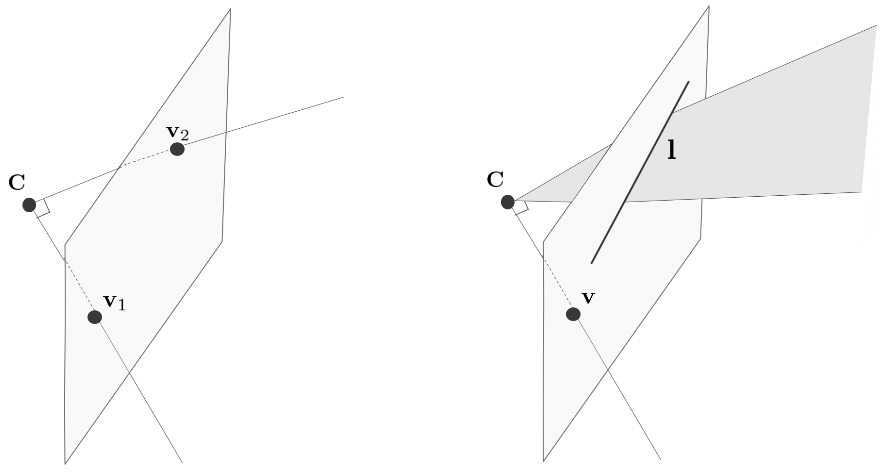
$$\mathbf{l} = \text{image of } \pi \cap \pi_\infty \tag{185}$$

Ultimately, **the vanishing line l represents the intersection of a plane π', parallel to the checkerboard plane π and containing the camera center, with the image plane.** π' can be computed by back-projecting the vanishing line $\mathbf{l}$. Depending on the relationship between the plane and line,

$$\pi' = \mathbf{K}^\intercal \mathbf{l} \tag{186}$$

can be calculated as such.

Orthogonality relationships amongst vanishing points and lines



The condition for lines back-projected from two vanishing points $\mathbf{v}_1, \mathbf{v}_2$ on the image plane to be orthogonal is as follows.

$$\mathbf{v}_1^\intercal \mathbf{w} \mathbf{v}_2 = 0 \tag{187}$$

In this case, $\mathbf{w} = \mathbf{K}^{-\intercal}\mathbf{K}^{-1}$ is the Image of Absolute Conic.

The condition for the line back-projected from the vanishing point $\mathbf{v}$ and the plane back-projected from the vanishing line $\mathbf{l}$ to be orthogonal is as follows.

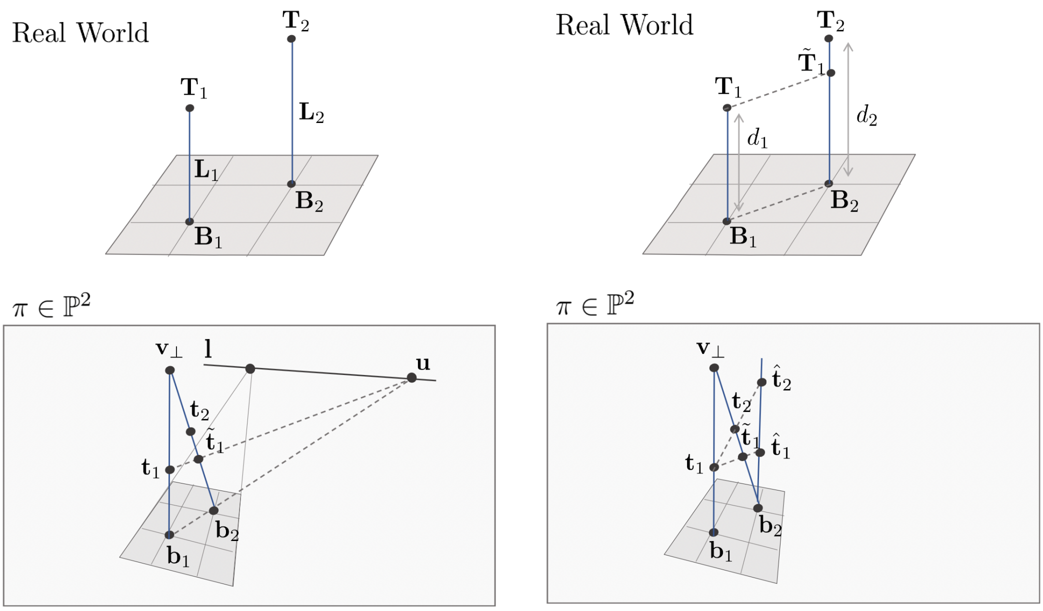
$$\mathbf{l} = \mathbf{w}\mathbf{v} \tag{188}$$

The condition for the planes back-projected from two lines $\mathbf{l}_1, \mathbf{l}_2$ on the image plane to be orthogonal to each other is as follows.

$$\mathbf{l}_1^\intercal \mathbf{w}^* \mathbf{l}_2 = 0 \tag{189}$$

In this case, $\mathbf{w}^*$ is the Image of Dual Absolute Conic.

Affine 3D measurements and reconstruction



The vanishing point perpendicular to the 3D space plane π is called the vertical vanishing point $\mathbf{v}_\perp$.

Result 8.24

Using the vanishing line $\mathbf{l}$ and the vertical vanishing point, the size of objects can be calculated up to a scale parameter. To be precise, **if you know the vanishing line $\mathbf{l}$ and the vertical vanishing point $\mathbf{v}_\perp$, you can determine the relative lengths of line segments perpendicular to the plane π.**

For example, suppose there are two points $\mathbf{B}_1, \mathbf{B}_2$ located on the 3D space plane π and lines $\mathbf{L}_1, \mathbf{L}_2$ passing through them and perpendicular to π. If there are endpoints $\mathbf{T}_1, \mathbf{T}_2$ of lines $\mathbf{L}_1, \mathbf{L}_2$, then **the relative lengths of $\mathbf{T}_1, \mathbf{T}_2$ can be measured using the vanishing line $\mathbf{l}$ and the vertical vanishing point $\mathbf{v}_\perp$.**

First, all elements are projected onto the image plane. At this time, define the point where the line connecting $\mathbf{b}_1, \mathbf{b}_2$ meets the vanishing line $\mathbf{l}$ as $\mathbf{u}$. Also, if you draw a line parallel to $\overline{\mathbf{b}_1\mathbf{b}_2}$ from $\mathbf{t}_1$, this line will touch $\mathbf{u}$. At this time, the intersection of $\overline{\mathbf{t}_1\mathbf{u}}$ and $\overline{\mathbf{v}_\perp\mathbf{b}_2}$ is defined as $\tilde{\mathbf{t}}_1$.

Next, project the line $\overline{\mathbf{v}_\perp\mathbf{b}_2}$ onto the $\mathbb{P}^1$ space where the vertical vanishing point $\mathbf{v}_\perp$ becomes the infinite point $(1, 0)$ and $\mathbf{b}_1$ becomes the origin $(0, 1)$. The Homography $\mathbf{H}_{2\times2}$ used at this time is as follows.

$$\mathbf{H}_{2\times2} = \begin{bmatrix} 1 & 0 \\ 1 & -\mathbf{v}_\perp \end{bmatrix} \tag{190}$$

$\mathbf{H}_{2\times2}$ preserves the Cross-Ratio. Next, using the ratio of lengths, calculate the ratio $d_1 : d_2 = \mathbf{b}_1\tilde{\mathbf{t}}_1 : \mathbf{b}_1\mathbf{t}_2$.

$$\frac{d_1}{d_2} = \frac{\tilde{\mathbf{t}}_1(\mathbf{v}_\perp - \mathbf{t}_2)}{\mathbf{t}_2(\mathbf{v}_\perp - \tilde{\mathbf{t}}_1)} \tag{191}$$

If the vertical vanishing point $\mathbf{v}_\perp$ and the camera's principal axis are perpendicular in the image plane, the vertical vanishing point does not intersect, and the ratio can be calculated simply as follows.

$$\frac{d_1}{d_2} = \frac{\tilde{\mathbf{t}}_1 - \mathbf{b}_2}{\mathbf{t}_2 - \mathbf{b}_2} \tag{192}$$

Determining camera calibration K from a single view

In a single-view image, two constraints are needed to determine the internal parameters $\mathbf{K}$. **These are the image constraint and the internal parameter constraint**.

The image constraint includes two orthogonal vanishing points $\mathbf{v}_1, \mathbf{v}_2$ on the image plane, such that $\mathbf{v}_1^\intercal \mathbf{w} \mathbf{v}_2 = 0$, and the case where the vanishing line $\mathbf{l}$ and the vanishing point $\mathbf{v}$ are orthogonal, for which their back-projections are $\mathbf{l} = \mathbf{w}\mathbf{x}$.

$$\begin{aligned} \mathbf{v}_1^\intercal \mathbf{w} \mathbf{v}_2 &= 0 \\ \mathbf{l} \times (\mathbf{w}\mathbf{v}) &= 0 \end{aligned} \tag{193}$$

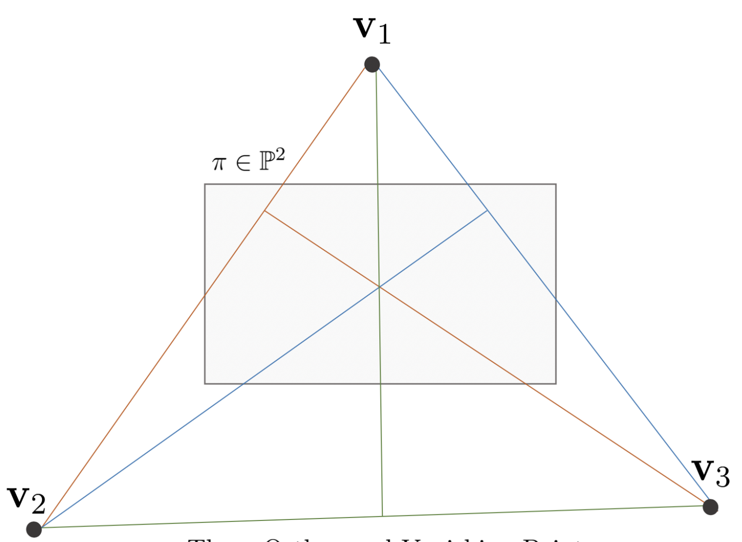
Result 8.26

The internal parameter constraint includes the case where $\mathbf{K}$ has Zero-Skew, where $w_{12} = w_{21} = 0$, and also, where it has Square Pixels, where $w_{12} = w_{21} = 0, w_{11} = w_{22}$.

$$\begin{aligned} & w_{12} = w_{21} \quad \text{for zero skew} \\ & w_{12} = w_{21} = 0, w_{11} = w_{22} \quad \text{for square pixel} \end{aligned} \tag{194}$$

As such, $\mathbf{w} = \begin{bmatrix} w_1 & w_2 & w_4 \\ w_2 & w_3 & w_5 \\ w_4 & w_5 & w_6 \end{bmatrix}$ parameters can be found after securing sufficient constraints and $\mathbf{w}$ is aligned as a 6-dimensional vector to form a $\mathbf{A}\mathbf{x} = 0$ linear system. Next, **use Singular Value Decomposition (SVD) to calculate the value of $\mathbf{w} = \mathbf{K}^{-\intercal}\mathbf{K}^{-1}$, then use Cholesky Decomposition to decompose $\mathbf{K}^{-\intercal}\mathbf{K}^{-1}$**. This allows the calculation of the internal parameters $\mathbf{K}$.

Calibration from three orthogonal vanishing points



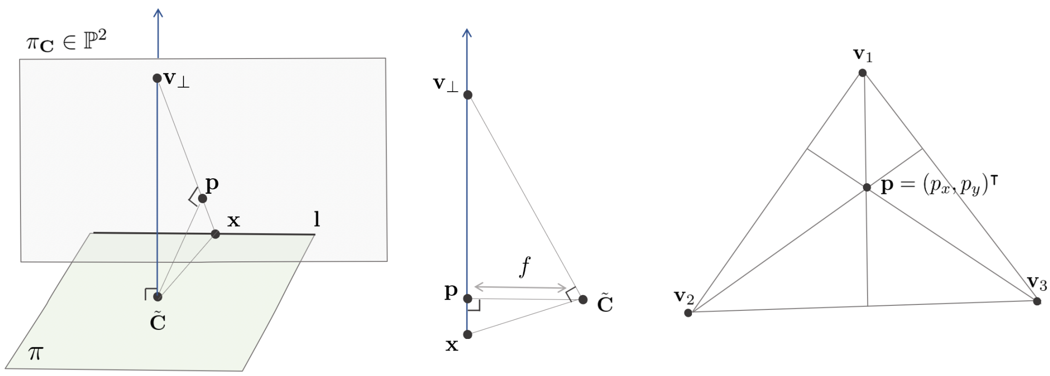
Three Orthogonal Vanishing Points

Using three mutually orthogonal vanishing points $\mathbf{v}_i$, $i = 1, 2, 3$, the internal parameters $\mathbf{K}$ can be calculated when $\mathbf{K}$ has Zero-Skew and Square Pixel. First, since $\mathbf{K}$ has Zero-Skew and Square Pixel, the following holds.

$$\mathbf{w} = \begin{bmatrix} w_1 & 0 & w_2 \\ 0 & w_1 & w_3 \\ w_2 & w_3 & w_4 \end{bmatrix} \tag{195}$$

Having a total of 4 degrees of freedom, $\mathbf{K}$ can be determined using three orthogonal vanishing points. First, use the orthogonal characteristics of each vanishing point to calculate $\mathbf{v}_i^\intercal \mathbf{w} \mathbf{v}_j$, $^\forall i \neq j$. Then, convert $\mathbf{w}$ into a vectorized form to form a $\mathbf{Ax} = 0$ linear system, and by calculating Nul $(\mathbf{A})$, find $\mathbf{w}$. Next, use Cholesky Decomposition to decompose $\mathbf{w} = \mathbf{K}^{-\intercal}\mathbf{K}^{-1}$. This way, the internal parameters $\mathbf{K}$ can be calculated.

Computation of focal length and principal point using vanishing point and vanishing line

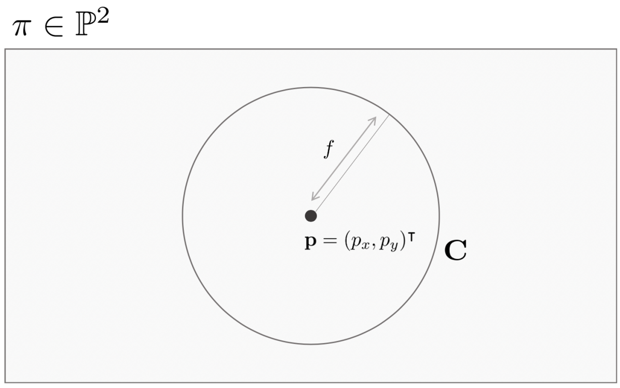


If there is a vanishing line $\mathbf{l}$ that can be obtained from the 3D space plane π and a vertical vanishing point $\mathbf{v}_\perp$ perpendicular to π, these can be used to determine the focal length f and the principal point.

The method for determining the focal length f is as follows. If the vertical vanishing point is $\mathbf{v}_\perp$ and the intersection line of the image plane $\pi_\mathbf{C}$ and the plane π is $\mathbf{l}$, then $\mathbf{v}_\perp\tilde{\mathbf{C}}$ and $\mathbf{x}\tilde{\mathbf{C}}$ are orthogonal to each other. If there is a point P where a perpendicular is dropped from the camera center to the image plane, then the length of $\tilde{\mathbf{C}}\mathbf{p}$ is the focal length f. **Draw a circle with diameter $\mathbf{v}_\perp x$, and if the line drawn horizontally from point p intersects the circle at points $\mathbf{a}, \mathbf{b}$, one of these points is the camera center $\tilde{\mathbf{C}}$, and this equals $\mathbf{ap} = \mathbf{bp}$, which is the focal length f.**

The method for determining the principal point P is as follows. If the vertical vanishing point is $\mathbf{v}_1$ and the intersection line of the image plane and the plane π is $\mathbf{l}_1$, then dropping a perpendicular from $\mathbf{v}_1$ to $\mathbf{l}_1$ positions the principal point $\mathbf{p}$ on this perpendicular. Because of this feature, **if there are three different vertical vanishing points, their orthocenter becomes the principal point p**.

The calibrating conic



$$\mathbf{C} = \mathbf{K}^{-\intercal}\mathbf{D}\mathbf{K}^{-1} = \begin{bmatrix} 1 & & \\ & 1 & \\ & & -f^2 \end{bmatrix}$$

$$\text{where, } \mathbf{K} = \begin{bmatrix} f & & \\ & f & \\ & & 1 \end{bmatrix}$$

The IAC (image of absolute conic) is a useful tool that can measure the angles between lines back-projected from points on the image plane and perform Metric Rectification, but it has the disadvantage of being unvisualizable due to the nature of Circular Points not having real roots. To compensate for this, the Calibration Conic was devised. The Calibration Conic means the Image Conic projected from the Cone $X^2+Y^2=Z^2$ and has the advantage of being visualizable.

When there is a camera projection $\mathbf{P}=\mathbf{K}[\mathbf{I}|0]$, points on the Calibration Conic are projected as follows.

$$\mathbf{C}=\mathbf{K}^{-\intercal}\begin{bmatrix}1 & & \\ & 1 & \\ & & -1\end{bmatrix}\mathbf{K}^{-1} \tag{196}$$

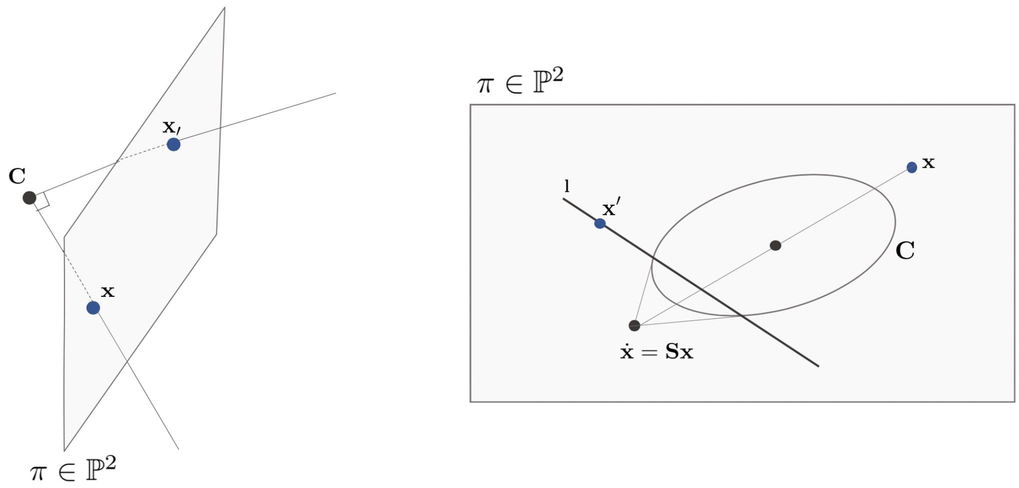
If $\mathbf{D}=\begin{bmatrix}1 & & \\ & 1 & \\ & & -1\end{bmatrix}$, then $\mathbf{C}=\mathbf{K}^{-\intercal}\mathbf{D}\mathbf{K}^{-1}$ can be represented as such. If $\mathbf{K}=\begin{bmatrix}f & & \\ & f & \\ & & 1\end{bmatrix}$, then the Calibration Conic can be represented as

$$\mathbf{C}=\begin{bmatrix}1 & & \\ & 1 & \\ & & -f^2\end{bmatrix} \tag{197}$$

and in this case, **the Calibration Conic represents a circle on the image with the principal point as the origin and the radius as the focal length f**.

$\mathbf{C}$ can be redefined as follows.

$$\begin{aligned}\mathbf{C}&=\mathbf{K}^{-\intercal}\mathbf{D}\mathbf{K}^{-1}=\mathbf{K}^{-\intercal}\mathbf{K}^{-1}\mathbf{K}\mathbf{D}\mathbf{K}^{-1}\\ \mathbf{C}&=\mathbf{w}\mathbf{S}\quad \text{where, } \mathbf{S}=\mathbf{K}\mathbf{D}\mathbf{K}^{-1}\end{aligned} \tag{198}$$



For any point $\mathbf{x}=\mathbf{K}\tilde{\mathbf{X}}$ on the image, $\mathbf{S}\mathbf{x}$ represents the reflected point by the Calibration Conic.

$$\begin{aligned}\mathbf{S}\mathbf{x}&=\mathbf{K}\mathbf{D}\mathbf{K}^{-1}\mathbf{K}\tilde{\mathbf{X}}\\ &=\mathbf{K}\mathbf{D}\begin{bmatrix}X\\Y\\Z\end{bmatrix}=\mathbf{K}\begin{bmatrix}X\\Y\\-Z\end{bmatrix}\end{aligned} \tag{199}$$

For two points $\mathbf{x},\mathbf{x}'$ on the image plane, if the lines back-projected from these two points are orthogonal to each other, $\mathbf{x}'^{\intercal}\mathbf{w}\mathbf{x}=0$ holds. Rewriting this equation,

$$\begin{aligned}&\mathbf{x}'^{\intercal}\mathbf{w}\mathbf{x}=\mathbf{x}'^{\intercal}\mathbf{C}\mathbf{S}^{-1}\mathbf{x}=\mathbf{x}'^{\intercal}\mathbf{C}\mathbf{S}\mathbf{x}=\mathbf{x}'^{\intercal}\mathbf{C}\dot{\mathbf{x}}\\ &\because \mathbf{S}^{-1}=\mathbf{S}\end{aligned} \tag{200}$$

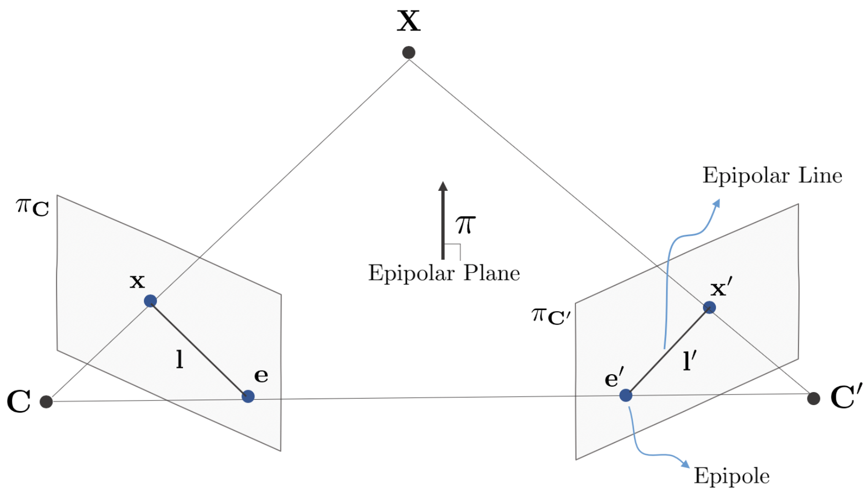
Here, $\dot{\mathbf{x}}=\mathbf{S}\mathbf{x}$ represents the point reflected by the Calibration Conic.

Result 8.30

Conclusively, **the line $\mathbf{C}\dot{\mathbf{x}}$ becomes the line connecting the reflected point $\dot{\mathbf{x}}$ and the tangents of the Calibration Conic, and the point $\mathbf{x}'$ exists on the line $\mathbf{C}\dot{\mathbf{x}}$**.

6 Epipolar Geometry and the Fundamental Matrix

Epipolar geometry



Epipolar geometry refers to the **geometric relationship defined between two camera images**. It is independent of the structure of the 3D object and **depends solely on the internal parameters of the cameras and the relative pose between the two cameras**. As shown in the figure above, given the centers of two cameras $\mathbf{C}, \mathbf{C}'$ and a point $\mathbf{X}$ in three-dimensional space, the plane uniquely determined through these three points is called the **Epipolar Plane** π. Additionally, the point $\mathbf{P}'\mathbf{C} = \mathbf{e}'$, which is the projection of the camera center $\mathbf{C}$ through $\mathbf{P}'$, is referred to as the **Epipole** $\mathbf{e}'$ on the image plane $\pi_{\mathbf{C}'}$. The line connecting $\mathbf{e}'$ and $\mathbf{x}'$ is called the **Epipolar Line** $\mathbf{l}'$. Similarly, $\mathbf{e}, \mathbf{l}$ refer to the Epipole and Epipolar Line on the image plane $\pi_{\mathbf{C}}$ respectively.

The fundamental matrix F

The Fundamental Matrix is a **matrix $\mathbf{F} \in \mathbb{R}^{3\times3}$ with rank 2 given the centers of two cameras $\mathbf{C}, \mathbf{C}'$**. The matrix $\mathbf{F}$ satisfies the following for corresponding point pairs $\mathbf{x}, \mathbf{x}'$ on the image planes of the two cameras:

$$\mathbf{x}'^{\intercal}\mathbf{F}\mathbf{x} = 0 \tag{201}$$

The meaning varies depending on the order of multiplication. If $\mathbf{x}'^{\intercal}\mathbf{F}\mathbf{x} = 0$, then it is referred to as the Fundamental Matrix between the two cameras $\mathbf{C}, \mathbf{C}'$, and if $\mathbf{x}^{\intercal}\mathbf{F}'\mathbf{x}' = 0$, it is the Fundamental Matrix between $\mathbf{C}', \mathbf{C}$. Here, $\mathbf{F}^{\intercal} = \mathbf{F}'$ holds true.

From a geometric perspective, in $\mathbf{x}'^{\intercal}\mathbf{F}\mathbf{x} = 0$, **$\mathbf{F}\mathbf{x}$ signifies the Epipolar Line $\mathbf{l}' \in \pi_{\mathbf{C}'}$ corresponding to the point $\mathbf{x}$ on the image plane $\pi_{\mathbf{C}}$**. Therefore, **$\mathbf{F}$ can be considered as a function mapping the point $\mathbf{x}$ on the image plane $\pi_{\mathbf{C}}$ to the Epipolar Line $\mathbf{l}'$**.

$$\begin{aligned} &\mathbf{F}: \ \mathbf{x} \mapsto \mathbf{l}' \\ &\text{where,} \quad \mathbf{x} \in \pi_{\mathbf{C}} \in \mathbb{P}^2 \\ &\mathbf{l}' \in (\mathbb{P}^2)^{\vee} \end{aligned} \tag{202}$$

Proof

($\Rightarrow$) If $\mathbf{l}' = \mathbf{F}\mathbf{x}$, then for a point $\mathbf{x}'$ on the Epipolar Line $\mathbf{l}'$, the following holds:

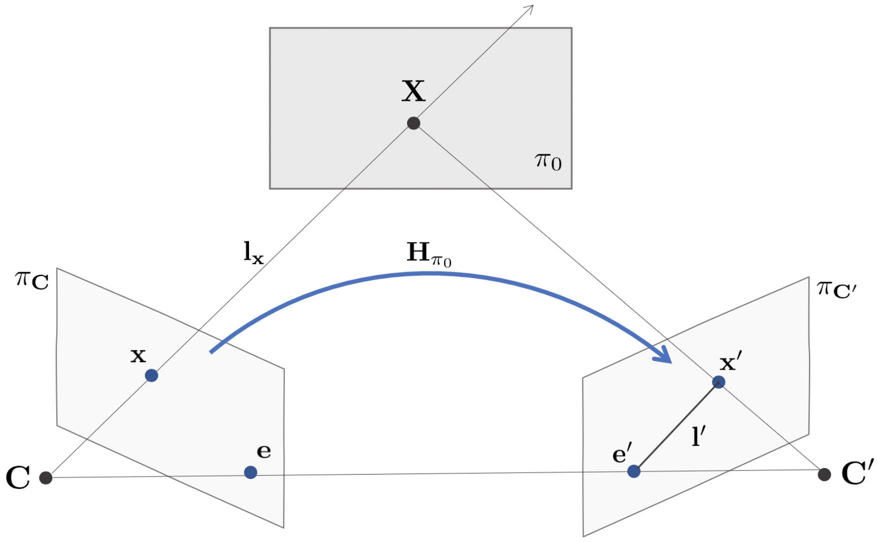
$$\mathbf{x}'^{\mathsf{T}}\mathbf{l}' = \mathbf{x}'^{\mathsf{T}}\mathbf{F}\mathbf{x} = 0 \quad {}^{\forall}\mathbf{x} \leftrightarrow \mathbf{x}' \tag{203}$$

($\Leftarrow$) Assuming that points $\mathbf{X}_1, \mathbf{X}_2$ are created through Back-projection of the point $\mathbf{x}$ on the image plane $\pi_{\mathbf{C}}$, and projecting these onto $\pi_{\mathbf{C}'}$ results in the points $\mathbf{x}'_1 = \mathbf{P}'\mathbf{X}_1$, $\mathbf{x}'_2 = \mathbf{P}'\mathbf{X}_2$. Then, due to the Fundamental Matrix $\mathbf{F}$, the following relations hold:

$$\begin{aligned} \mathbf{x}_1'^{\mathsf{T}}\mathbf{F}\mathbf{x} = 0 \\ \mathbf{x}_2'^{\mathsf{T}}\mathbf{F}\mathbf{x} = 0 \end{aligned} \tag{204}$$

Thus, $\mathbf{F}\mathbf{x}$ signifies the Epipolar Line $\mathbf{l}'$ orthogonal to $\mathbf{x}_i'^{\mathsf{T}}$, $i = 1, 2$.

Geometric derivation



When considering two cameras $\mathbf{C}, \mathbf{C}'$, let us assume that the line $\mathbf{l_x}$, Back-projected from the point $\mathbf{x}$ on the image plane $\pi_{\mathbf{C}}$, intersects a random plane π_0 at a point in three-dimensional space. If we consider the Homography $\mathbf{H}_{\pi_0}$ that connects the image points of $\mathbf{C}, \mathbf{C}'$,

$$\begin{aligned} \mathbf{H}_{\pi_0} : & \pi_{\mathbf{C}} \mapsto \pi_{\mathbf{C}'} \\ & \mathbf{x} \mapsto \mathbf{x}' \\ & \mathbf{x} \mapsto \mathbf{P}'(\mathbf{l_x} \cap \pi_0) \end{aligned} \tag{205}$$

If the point projected onto $\pi_{\mathbf{C}'}$ is $\mathbf{x}' = \mathbf{P}'(\mathbf{l_x} \cap \pi_0)$, and we define the Epipolar Line $\mathbf{l}'$ as the line connecting $\mathbf{x}'$ and the Epipole $\mathbf{e}'$,

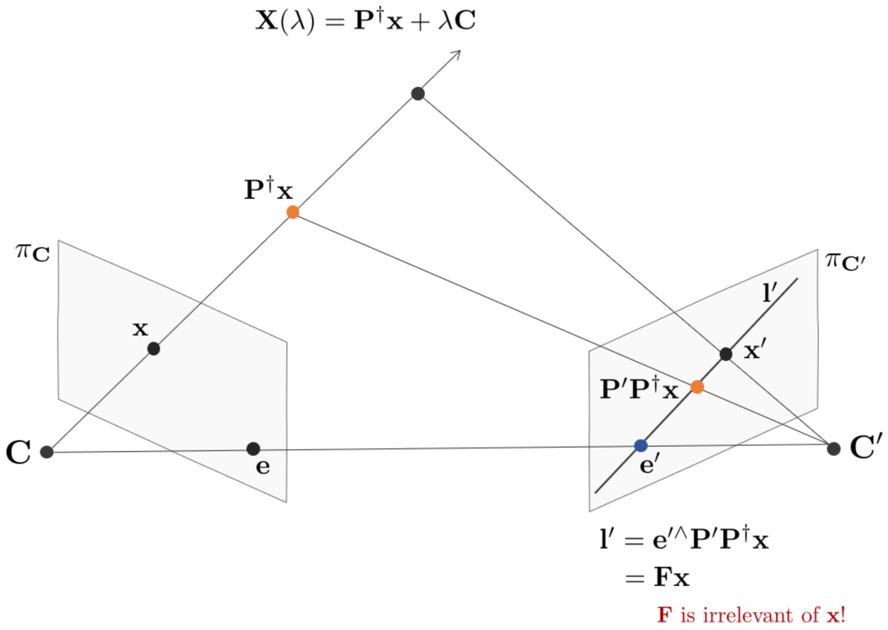
$$\begin{aligned} \mathbf{e}' \times \mathbf{x}' &= \mathbf{e}'^{\wedge}\mathbf{x}' \\ &= \mathbf{e}'^{\wedge}\mathbf{H}_{\pi_0}(\mathbf{x}) \quad \text{is Epipolar Line.} \end{aligned} \tag{206}$$

In this case, using the formula $\mathbf{l}' = \mathbf{F}\mathbf{x}$,

$$\begin{aligned} \mathbf{l}' &= \mathbf{F}\mathbf{x} \\ &= \mathbf{e}'^{\wedge}\mathbf{H}_{\pi_0}\mathbf{x} \end{aligned} \tag{207}$$

Thus, $\mathbf{F} = \mathbf{e}'^{\wedge}\mathbf{H}_{\pi_0}$ is established. Here, $\mathbf{H}_{\pi_0}$ is a rank 3 matrix and $\mathbf{e}'^{\wedge}$ is a rank 2 matrix, so $\mathbf{F}$ is a rank 2 matrix.

Algebraic derivation



The Epipolar Line $\mathbf{l}' = \mathbf{P}'(\mathbf{X}(\lambda))$. Here, $\mathbf{X}(\lambda)$ represents the Back-projection line of $\mathbf{x}$ passing through the center of camera $\mathbf{C}$. $\mathbf{X}(\lambda)$ can be rewritten as follows.

$$\mathbf{X}(\lambda) = \mathbf{P}^{\dagger}\mathbf{x} + \lambda\mathbf{C} \quad \lambda \in \mathbb{R} \tag{208}$$

Here, $\mathbf{P}^{\dagger}$ represents the Pseudo Inverse of $\mathbf{P}$ and

$$\mathbf{P}^{\dagger} = \mathbf{P}^{\intercal}(\mathbf{P}\mathbf{P}^{\intercal})^{-1} \tag{209}$$

Then, $\mathbf{l}' = \mathbf{P}'(\mathbf{X}(\lambda))$ is as follows.

$$\begin{aligned} \mathbf{P}'(\mathbf{X}(\lambda)) = \mathbf{P}'\mathbf{P}^{\dagger}\mathbf{x} + \lambda\mathbf{P}'\mathbf{C} \\ \mathbf{P}'\mathbf{P}^{\dagger}\mathbf{x} + \lambda\mathbf{e}' \end{aligned} \tag{210}$$

$\mathbf{P}'\mathbf{C}$ signifies the point projected onto $\pi_{\mathbf{C}'}$ from the center of camera $\mathbf{C}$, thus becoming the Epipole $\mathbf{e}'$. When $\lambda = 0$, $\mathbf{P}'(\mathbf{X}(0)) = \mathbf{P}'\mathbf{P}^{\dagger}\mathbf{x}$ and when $\lambda = \infty$, $\mathbf{P}'(\mathbf{X}(\infty)) = \mathbf{P}'\mathbf{C} = \mathbf{e}'$. Therefore, the Epipolar Line $\mathbf{l}'$ connects these two points, so

$$\begin{aligned} \mathbf{l}' &= \mathbf{e}'^{\wedge}\mathbf{P}'\mathbf{P}^{\dagger}\mathbf{x} \\ &= \mathbf{F}\mathbf{x} \\ \therefore \mathbf{F} &= \mathbf{e}'^{\wedge}\mathbf{P}'\mathbf{P}^{\dagger} \end{aligned} \tag{211}$$

Additionally, since Epipole $\mathbf{e}'$ is included in the Epipolar Line $\mathbf{l}'$ and holds for all $\mathbf{x}_i$,

$$\mathbf{e}'^{\intercal}\mathbf{F}\mathbf{x}_i = (\mathbf{e}'^{\intercal}\mathbf{F})\mathbf{x}_i = 0 \quad {}^{\forall}\mathbf{x}_i \tag{212}$$

$\mathbf{e}'^{\intercal}\mathbf{F} = 0$ holds true. **In conclusion, $\mathbf{e}'$ is the Left null vector of $\mathbf{F}$**. Similarly, $\mathbf{e}$ becomes the (right) null vector of $\mathbf{F}$.

Properties of the fundamental matrix

The properties of the Fundamental Matrix $\mathbf{F}$ are as follows:

- **Transpose:** If $\mathbf{F}$ is the Fundamental Matrix for two cameras $(\mathbf{P}, \mathbf{P}')$, then $\mathbf{F}^{\intercal}$ becomes the Fundamental Matrix for $(\mathbf{P}', \mathbf{P})$.

- **Epipolar Lines:** For a corresponding point $\mathbf{x}$ in the first image, the corresponding Epipolar Line in the second image can be expressed as $\mathbf{l}' = \mathbf{F}\mathbf{x}$. Similarly, for a corresponding point $\mathbf{x}'$ in the second image, the corresponding Epipolar Line in the first image can be expressed as $\mathbf{l} = \mathbf{F}^\intercal \mathbf{x}'$.
- **The Epipole:** Epipolar Line $\mathbf{l}' = \mathbf{F}\mathbf{x}$ always passes through $\mathbf{e}'$ for any point $\mathbf{x}$ other than $\mathbf{e}$. Therefore, $\mathbf{e}'$ satisfies $\mathbf{e}'^\intercal(\mathbf{F}\mathbf{x}) = (\mathbf{e}'^\intercal \mathbf{F})\mathbf{x} = 0$ for all $\mathbf{x}$, indicating that $\mathbf{e}'^\intercal \mathbf{F} = 0$. Hence, $\mathbf{e}'$ is the Left null-vector of $\mathbf{F}$. Similarly, $\mathbf{F}\mathbf{e} = 0$ means that $\mathbf{e}$ is the Right null-vector of $\mathbf{F}$.
- **Fundamental Matrix is a rank 2 Homogeneous matrix with 7 degrees of freedom (DOF) and lacks an inverse matrix.** As a 3x3 matrix, it loses one degree of freedom due to the Scale Ambiguity in Homogeneous coordinates, and an additional degree of freedom is lost due to the constraint $\det \mathbf{F} = 0$, resulting in 7 DOF.

The Epipolar Line Homography

Let there be two cameras $\mathbf{C}, \mathbf{C}'$ and let's denote a point on their image planes as $\mathbf{x}, \mathbf{x}'$ respectively. Corresponding to these, there exist Epipolar Lines $\mathbf{l}, \mathbf{l}'$. There is a specific relationship between $\mathbf{l}$ and $\mathbf{l}'$ which is represented by

$$\mathbb{P}(\pi)^\vee \mapsto \mathbb{P}(\pi')^\vee \tag{213}$$

This relationship is provided by the **Homography matrix**.

Result 9.5

When there exists a line $\mathbf{k}$ that passes through the camera center $\mathbf{C}$ but does not pass through the Epipole $\mathbf{e}$, $\mathbf{k}$ will necessarily meet the Epipolar Line $\mathbf{l}$ at a point $\mathbf{p}$. $\mathbf{p}$ can be calculated as

$$\mathbf{p} = \mathbf{k}^\wedge \mathbf{l} \tag{214}$$

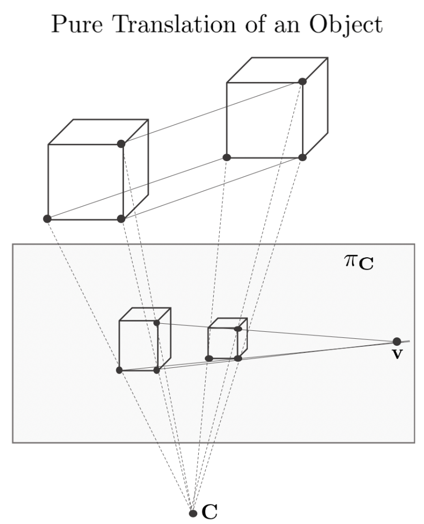
and since $\mathbf{p}$ is projected onto Epipolar Line $\mathbf{l}'$ by the Fundamental Matrix, the following relation holds:

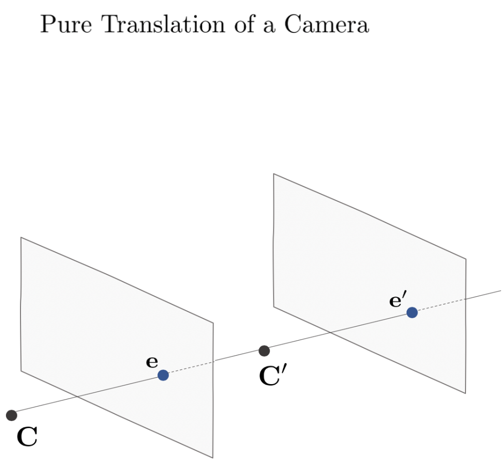
$$\begin{aligned} \mathbf{l}' &= \mathbf{F}\mathbf{k}^\wedge \mathbf{l} \\ &= \mathbf{H}\mathbf{l} \end{aligned} \tag{215}$$

Thus, **the relationship between $\mathbf{l}, \mathbf{l}'$ is established by the Homography matrix $\mathbf{H} = \mathbf{F}\mathbf{k}^\wedge$**.

Fundamental matrices arising from special motions

Pure translation





Pure translation implies moving the camera center without any rotation. In this scenario, moving the camera while fixing all objects in the world is equivalent to fixing the camera and moving all objects in the world. If we denote the initial camera matrix and camera center as $\mathbf{P}, \mathbf{C}$ respectively, and the camera matrix and camera center after pure translation as $\mathbf{P}', \mathbf{C}'$ respectively, **the length of the baseline between the two cameras is equal to the amount of pure translation of the camera center**.

Also, **if the Epipoles of both cameras are denoted as e, e' respectively, the position of both Epipoles is the same, and this is referred to as the vanishing point**.

$$\mathbf{e} = \mathbf{e}' = \mathbf{v} \tag{216}$$

In such a pure translation situation, $\mathbf{e} = \mathbf{e}'$ is referred to as Auto Epipolar.

Algebraically deriving the Fundamental Matrix, if the initial camera matrix is $\mathbf{P} = \mathbf{K}[\mathbf{I}|0]$ and after pure translation the camera matrix is $\mathbf{P}' = \mathbf{K}[\mathbf{I}|\mathbf{t}]$, then the Fundamental Matrix $\mathbf{F}$ from the previous section is $\mathbf{F} = \mathbf{e}'^{\wedge}\mathbf{P}'\mathbf{P}^{\dagger}$ therefore,

$$\begin{aligned} \mathbf{F} &= \mathbf{e}'^{\wedge}\mathbf{P}'\mathbf{P}^{\dagger} \\ &= \mathbf{e}'^{\wedge}\mathbf{K}\mathbf{K}^{-1} \\ &= \mathbf{e}'^{\wedge} \end{aligned} \tag{217}$$

Therefore, **in a situation of pure translation, the Fundamental Matrix is $\mathbf{F} = \mathbf{e}'^{\wedge}$**.

Retrieving the camera matrices

Projective invariance and canonical cameras

Result 9.8

When there exist points $\mathbf{x}, \mathbf{x}'$ on the image planes corresponding to two cameras $\mathbf{C}, \mathbf{C}'$ and a Fundamental Matrix $\mathbf{F}$, $\mathbf{F}$ remains the same regardless of the Homography transformation of the $\mathbf{x} \leftrightarrow \mathbf{x}'$ pair. In other words, when there exists a Homography transformation $\mathbf{H} \in \mathbb{R}^{4\times4}$ that satisfies $\mathbb{P}^3 \mapsto \mathbb{P}^3$, $\mathbf{H}$ satisfies

$$(\mathbf{P}, \mathbf{P}') \mapsto (\mathbf{PH}, \mathbf{P}'\mathbf{H}) \tag{218}$$

and **irrespective of such $\mathbf{H}$, the Fundamental Matrix $\mathbf{F}$ remains the same**.

Proof

For the corresponding point pairs $(\mathbf{PH}, \mathbf{P}'\mathbf{H})$,

$$\mathbf{x}'^{\intercal}\tilde{\mathbf{F}}\mathbf{x} = 0 \tag{219}$$

holds. Since $\mathbf{x} = \mathbf{PHX}, \mathbf{x}' = \mathbf{P}'\mathbf{HX}$, substituting this into the above equation,

$$\mathbf{X}^{\intercal}\mathbf{H}^{\intercal}\mathbf{P}'^{\intercal}\tilde{\mathbf{F}}\mathbf{PHX} = 0 \tag{220}$$

holds for all ${}^{\forall}\mathbf{X} \in \mathbb{P}^3$. Replacing $\mathbf{X} = \mathbf{H}^{-1}(\mathbf{HX}), \tilde{\mathbf{X}} = \mathbf{HX}$ in the above equation, we get

$$\begin{aligned} &\tilde{\mathbf{X}}^{\intercal}\mathbf{H}^{-\intercal}\mathbf{H}^{\intercal}\mathbf{P}'^{\intercal}\tilde{\mathbf{F}}\mathbf{PHH}^{-1}\tilde{\mathbf{X}} = 0 \\ &\tilde{\mathbf{X}}^{\intercal}\mathbf{P}'^{\intercal}\tilde{\mathbf{F}}\mathbf{P}\tilde{\mathbf{X}} = 0 \\ &\tilde{\mathbf{x}}'^{\intercal}\tilde{\mathbf{F}}\tilde{\mathbf{x}} = 0 \end{aligned} \tag{221}$$

Thus, **in conclusion, $\mathbf{F} = \tilde{\mathbf{F}}$ is valid**.

Canonical form of camera matrices

Following this property, multiple $(\mathbf{P}, \mathbf{P}')$ pairs correspond to the same Fundamental Matrix $\mathbf{F}$, hence $\mathbf{F}$ and $(\mathbf{P}, \mathbf{P}')$ have a One-to-Many Correspondence relationship. Therefore, despite this ambiguity, to accurately represent the transformation of $\mathbf{F}$, the initial camera matrix $\mathbf{P}$ is simply represented as Canonical Form $\mathbf{P} = [\mathbf{I}|0]$, $\mathbf{P}' = [\mathbf{M}|\mathbf{m}]$. However, stating $\mathbf{P} = [\mathbf{I}|0]$ does not uniquely determine $\mathbf{P}'$.

When any camera matrix $\mathbf{P} \in \mathbb{R}^{3\times4}$ is given, it can be changed to a matrix $\mathbf{P}^* = \begin{bmatrix} \mathbf{P} \\ \mathbf{r}_4 \end{bmatrix} \in \mathbb{R}^{4\times4}$ that has an inverse, then

$$\mathbf{P}^*\mathbf{H} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{bmatrix} \tag{222}$$

satisfies $\mathbf{PH} = [\mathbf{I}|\mathbf{0}]$. Therefore, **assuming the existence of H, any arbitrary camera matrix P can be written in Canonical Form.**

What then can the Fundamental Matrix $\mathbf{F}$ be written in terms of $\mathbf{M}, \mathbf{m}$?

Firstly, when there exist arbitrary camera matrices $\mathbf{P} = \mathbf{K}[\mathbf{I}|0], \mathbf{P}' = \mathbf{K}'[\mathbf{R}|\mathbf{t}]$, the following formulas hold.

$$\begin{aligned} &\mathbf{P}\mathbf{P}^\intercal = \mathbf{K}^2 \\ &\mathbf{P}^\dagger = \mathbf{P}^\intercal(\mathbf{P}\mathbf{P}^\intercal)^{-1} = \begin{bmatrix} \mathbf{K}^{-1} \\ 0 \end{bmatrix} \\ &\mathbf{C} = \begin{bmatrix} 0 & 1 \end{bmatrix}^\intercal \\ &\mathbf{e}' = \mathbf{P}'\mathbf{C} = \mathbf{K}'\mathbf{t} \end{aligned} \tag{223}$$

Therefore, the corresponding Fundamental Matrix $\mathbf{F}$ is as follows.

$$\begin{aligned} \mathbf{F} &= \mathbf{e}'^\wedge \mathbf{P}'\mathbf{P}^\dagger \\ &= (\mathbf{K}'\mathbf{t})^\wedge \mathbf{K}'[\mathbf{R}|\mathbf{t}]\mathbf{K}^{-1} \begin{bmatrix} \mathbf{I} \\ 0 \end{bmatrix} \\ &= (\mathbf{K}'\mathbf{t})^\wedge \mathbf{K}'\mathbf{R}\mathbf{K}^{-1} \end{aligned} \tag{224}$$

Result 9.9

Considering the Canonical Form, the Fundamental Matrix $\mathbf{F}$ for the two camera matrices $\mathbf{P} = [\mathbf{I}|0], \mathbf{P}' = [\mathbf{M}|\mathbf{m}]$ satisfies the following formula.

$$\mathbf{F} = \mathbf{m}^\wedge \mathbf{M} \tag{225}$$

The F of the Canonical Form is $\mathbf{F} = \mathbf{m}^\wedge \mathbf{M}$.

Projective Ambiguity of Cameras given F

Theorem 9.10

The Fundamental Matrix $\mathbf{F}$ is invariant under Homography transformation, leading to inherent ambiguity. If the $\mathbf{F}$ for camera matrix pairs $(\mathbf{P}, \mathbf{P}')$ and $(\tilde{\mathbf{P}}, \tilde{\mathbf{P}}')$ are the same, there exists a Homography $\mathbf{H}$ that connects $(\mathbf{P}, \mathbf{P}')$ and $(\tilde{\mathbf{P}}, \tilde{\mathbf{P}}')$.

$$\begin{aligned} &{}^\exists\mathbf{H} \in \mathrm{PGL}_4 \\ &(\tilde{\mathbf{P}}, \tilde{\mathbf{P}}') = (\mathbf{P}, \mathbf{P}')\mathbf{H} \end{aligned} \tag{226}$$

Proof

Given two pairs of camera matrices $(\mathbf{P}, \mathbf{P}')$ and $(\tilde{\mathbf{P}}, \tilde{\mathbf{P}}')$, when written in Canonical Form, they are as follows:

$$\begin{aligned} &\mathbf{P} = \tilde{\mathbf{P}} = [\mathbf{I}|0] \\ &\mathbf{P}' = [\mathbf{A}|\mathbf{a}] \\ &\tilde{\mathbf{P}}' = [\tilde{\mathbf{A}}|\tilde{\mathbf{a}}] \end{aligned} \tag{227}$$

Here, the Fundamental Matrix $\mathbf{F}$ is defined as:

$$\mathbf{F} = \mathbf{a}^\wedge \mathbf{A} = \tilde{\mathbf{a}}^\wedge \tilde{\mathbf{A}} \tag{228}$$

Due to the properties of $\mathbf{F}$, $\mathbf{a}^\intercal\mathbf{F} = 0$ and $\tilde{\mathbf{a}}^\intercal\mathbf{F} = 0$ hold true:

$$\begin{aligned} &\mathbf{a}^\intercal\mathbf{F} = \mathbf{a}^\intercal\mathbf{a}^\wedge\mathbf{A} = 0 \\ &\tilde{\mathbf{a}}^\intercal\mathbf{F} = \tilde{\mathbf{a}}^\intercal\tilde{\mathbf{a}}^\wedge\tilde{\mathbf{A}} = 0 \end{aligned} \tag{229}$$

Here, $\mathbf{a}$ and $\tilde{\mathbf{a}}$ form the rank 1 Left Null Space of $\mathbf{F}.Thus,$a$, arerelatedby:$

$$\tilde{\mathbf{a}} = k\mathbf{a}, \quad k \neq 0 \in \mathbb{R} \tag{230}$$

Substituting $\tilde{\mathbf{a}}$ and re-expressing it leads to:

$$\begin{aligned} \mathbf{F} &= \mathbf{a}^{\wedge}\mathbf{A} = \tilde{\mathbf{a}}^{\wedge}\tilde{\mathbf{A}} = k\mathbf{a}^{\wedge}\tilde{\mathbf{A}} = 0 \\ &= \mathbf{a}^{\wedge}(k\tilde{\mathbf{A}} - \mathbf{A}) = 0 \end{aligned} \tag{231}$$

Therefore, $(k\tilde{\mathbf{A}} - \mathbf{A})$ is parallel to $\mathbf{a}$. Hence, each column of the matrices are in a scalar relationship:

$$k\tilde{\mathbf{A}} - \mathbf{A} = \mathbf{a}\mathbf{v}^{\intercal} \quad \text{for some } \mathbf{v} \tag{232}$$

Thus, $\tilde{\mathbf{A}} = k^{-1}(\mathbf{A} + \mathbf{a}\mathbf{v}^{\intercal})$. The camera matrix pairs can be represented as:

$$\begin{aligned} \mathbf{P} &= \tilde{\mathbf{P}} = [\mathbf{I}|0] \\ \mathbf{P}' &= [\mathbf{A}|\mathbf{a}] \\ \tilde{\mathbf{P}}' &= [k^{-1}(\mathbf{A} + \mathbf{a}\mathbf{v}^{\intercal}) \mid k\mathbf{a}] \end{aligned} \tag{233}$$

In conclusion, if the matrix transforming camera matrix pair $(\mathbf{P}, \mathbf{P}')$ to Canonical Form is $\mathbf{H}_1$, and the matrix transforming $(\tilde{\mathbf{P}}, \tilde{\mathbf{P}}')$ to Canonical Form is $\mathbf{H}_2$, then

$$(\mathbf{P}, \mathbf{P}')\mathbf{H}_1\mathbf{H} = (\tilde{\mathbf{P}}, \tilde{\mathbf{P}}')\mathbf{H}_2 \tag{234}$$

a Homography H exists satisfying this, thereby making $(\mathbf{P}, \mathbf{P}')$ and $(\tilde{\mathbf{P}}, \tilde{\mathbf{P}}')$ projectively equivalent.

$$(\mathbf{P}, \mathbf{P}') \sim (\tilde{\mathbf{P}}, \tilde{\mathbf{P}}') \tag{235}$$

Canonical Cameras given F

Result 9.12

The necessary and sufficient condition for an arbitrary square matrix $\mathbf{F} \in \mathbb{R}^{3\times 3}$ to be the Fundamental Matrix corresponding to camera matrix pair $(\mathbf{P}, \mathbf{P}')$ is that $\mathbf{P}'^{\intercal}\mathbf{F}\mathbf{P}$ is a skew symmetric matrix.

$$\begin{aligned} &\mathbf{x}'\mathbf{F}\mathbf{x} = 0 \\ &\Leftrightarrow \mathbf{X}^{\intercal}\mathbf{P}'^{\intercal}\mathbf{F}\mathbf{P}\mathbf{X} = 0 \quad {}^{\forall}\mathbf{X} \\ &\Leftrightarrow \mathbf{P}'^{\intercal}\mathbf{F}\mathbf{P} \text{ is a skew symmetric matrix.} \end{aligned} \tag{236}$$

Result 9.13

Given an arbitrary square matrix $\mathbf{F} \in \mathbb{R}^{3\times 3}$ and an arbitrary skew symmetric matrix $\mathbf{S} \in \mathbb{R}^{3\times 3}$, if the camera matrix pair $(\mathbf{P}, \mathbf{P}')$ is as follows:

$$\begin{aligned} \mathbf{P} &= [\mathbf{I}|0] \\ \mathbf{P}' &= [\mathbf{S}\mathbf{F}|\mathbf{e}'] \quad \text{where, } \mathbf{e}' \text{ is epipole.} \end{aligned} \tag{237}$$

Then **$(\mathbf{P}, \mathbf{P}')$ possesses F as its Fundamental Matrix.**
To prove this, expand $\mathbf{P}'\mathbf{F}\mathbf{P}$:

$$\begin{aligned} \mathbf{P}'\mathbf{F}\mathbf{P} &= [\mathbf{S}\mathbf{F}|\mathbf{e}']^{\intercal}\mathbf{F}[\mathbf{I}|0] \\ &= \begin{bmatrix} \mathbf{F}^{\intercal}\mathbf{S}^{\intercal}\mathbf{F} & \mathbf{0} \\ \mathbf{e}'\mathbf{F} & 0 \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{F}^{\intercal}\mathbf{S}^{\intercal}\mathbf{F} & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix} \end{aligned} \tag{238}$$

Since it forms a skew symmetric matrix, this confirms $\mathbf{F}$ as the Fundamental Matrix.

The skew symmetric matrix $\mathbf{S}$ can be expressed in the form of a 3-dimensional vector $\mathbf{s}^{\wedge}$, and if $\mathbf{s}^{\intercal}\mathbf{e}' \neq 0$, then the rank of $\mathbf{P}' = [\mathbf{s}^{\wedge}\mathbf{F}|\mathbf{e}']$ is 3. For $\mathbf{P}'$ to have a rank of 3, the ranks of $\mathbf{s}^{\wedge}\mathbf{F}$ and $\mathbf{e}'$ must be 2 and 1, respectively. First, to prove $\mathbf{s}^\wedge\mathbf{F}$ has rank 2, since $\mathbf{e}'\mathbf{F} = 0$ satisfies the Fundamental Matrix property, the column space Col $\mathbf{F}$ and $\mathbf{e}'$ are orthogonal:

$$\text{Col } \mathbf{F} \perp \mathbf{e}' \tag{239}$$

Furthermore, since $\mathbf{s}^\wedge\mathbf{e}' \neq 0$, $\mathbf{s}$ does not reside within Col $\mathbf{F}$. The column space of $\mathbf{s}^\wedge\mathbf{F}$ is:

$$\begin{aligned} \text{Col } \mathbf{s}^\wedge\mathbf{F} &= \mathbf{s}^\wedge\text{Col } \mathbf{F} \\ &= \mathbf{s} \times \text{Col } \mathbf{F} \end{aligned} \tag{240}$$

And since Col F has rank 2, $\mathbf{s} \times$ Col F also has rank 2. Next, to prove $\mathbf{e}'$ is linearly independent from other column vectors, $\mathbf{s}^\intercal\text{Col } \mathbf{s}^\wedge\mathbf{F} = 0$ and $\mathbf{s}^\wedge\mathbf{e}' \neq 0$ therefore,

$$\mathbf{e}' \notin \text{Col } \mathbf{s}^\wedge\mathbf{F} \tag{241}$$

Conclusively, the rank of $\mathbf{P}' = [\mathbf{s}^\wedge\mathbf{F}|\mathbf{e}']$ is 3.

Result 9.14

When selecting $\mathbf{s}$ such that the rank of $\mathbf{P}'$ is 3, $\mathbf{s}$ that is not orthogonal to $\mathbf{e}'$ should be used, hence setting $\mathbf{s} = \mathbf{e}'$ leads to:

$$\begin{aligned} \mathbf{P} &= [\mathbf{I}|0] \\ \mathbf{P}' &= [\mathbf{e}'^\wedge\mathbf{F} \mid \mathbf{e}'] \end{aligned} \tag{242}$$

Thus, given the Fundamental Matrix F, computing the Left Null Space $\mathbf{e}'$ allows for the calculation of the camera pair $(\mathbf{P}, \mathbf{P}')$.

Using the proportionality defined earlier for the camera matrix pairs,

$$\begin{aligned} (\mathbf{P}, \mathbf{P}') &= ([\mathbf{I}|0], [\mathbf{A}|\mathbf{a}]) \\ (\tilde{\mathbf{P}}, \tilde{\mathbf{P}}') &= ([\mathbf{I}|0], [k^{-1}(\mathbf{A} + \mathbf{a}\mathbf{v}^\intercal) \mid k\mathbf{a}]) \end{aligned} \tag{243}$$

can be generalized as follows.

Result 9.15

$$\begin{aligned} (\mathbf{P}, \mathbf{P}') &= ([\mathbf{I}|0], [\mathbf{e}'\mathbf{F} \mid \mathbf{e}']) \\ (\tilde{\mathbf{P}}, \tilde{\mathbf{P}}') &= ([\mathbf{I}|0], [\mathbf{e}'\mathbf{F} + \mathbf{e}'\mathbf{v}^\intercal \mid \lambda\mathbf{e}']) \end{aligned} \tag{244}$$

The Null Space value $\mathbf{e}'$ being scale-invariant, using $\frac{1}{k}\mathbf{e}'$ results in the same outcome. Thus, it can be represented in the above form. The implication of this formula is that regardless of the arbitrary vector $\mathbf{e}'\mathbf{v}^\intercal$ and the arbitrary scale value λ multiplied with the existing $\mathbf{F}$, there is no change in the Fundamental Matrix. This is because substituting into $\mathbf{F} = \mathbf{e}'^\wedge\mathbf{P}'\mathbf{P}^\dagger = \mathbf{e}'^\wedge\tilde{\mathbf{P}}'\tilde{\mathbf{P}}^\dagger$ results in a consistent $\mathbf{F}$. **Thus, the above formula embodies the Projective Ambiguity of the Fundamental Matrix.**

The essential matrix

The Essential Matrix E refers to the Fundamental Matrix when the corresponding point pairs $\mathbf{x} \leftrightarrow \mathbf{x}'$ are in the normalized image coordinate system. Historically, the Essential Matrix was introduced by Longuet-Higgins before the Fundamental Matrix, and later, the Fundamental Matrix for uncalibrated corresponding point pairs was introduced as a generalized version of the Essential Matrix.

Normalized coordinates

Given any camera matrix $\mathbf{P} = \mathbf{K}[\mathbf{R}|\mathbf{t}]$, the point $\mathbf{x}$ on the image plane satisfies $\mathbf{x} = \mathbf{P}\mathbf{X} = \mathbf{K}[\mathbf{R}|\mathbf{t}]\mathbf{X}$. In this case, **the point on the normalized image plane is $\bar{\mathbf{x}}$**,

$$\bar{\mathbf{x}} = \mathbf{K}^{-1}\mathbf{x} \tag{245}$$

It holds that,

$$\mathbf{K}^{-1}\mathbf{P} = [\mathbf{R}|\mathbf{t}] \tag{246}$$

is referred to as the **normalized camera**. In a normalized camera, the camera matrix can be considered where $\mathbf{K} = \mathbf{I}$. The corresponding pair of normalized camera matrices $(\mathbf{P}, \mathbf{P}')$ is

$$(\mathbf{P}, \mathbf{P}') = ([\mathbf{I}|0],\ [\mathbf{R}|\mathbf{t}]) \tag{247}$$

thus in this case **the Fundamental Matrix $\mathbf{F} = \mathbf{t}^{\wedge}\mathbf{R}$ is specifically called the Essential Matrix $\mathbf{E}$**,

$$\mathbf{E} = \mathbf{t}^{\wedge}\mathbf{R} \tag{248}$$

Definition 9.16.

The Essential Matrix is a 3×3 square matrix, and $\mathbf{E} \in \mathbb{R}^{3\times3}$ is used to express the correlation between points in the normalized image coordinate system $\bar{\mathbf{x}} \leftrightarrow \bar{\mathbf{x}}'$.

$$\bar{\mathbf{x}}'^{\intercal}\mathbf{E}\bar{\mathbf{x}} = 0 \tag{249}$$

At this time, since $\bar{\mathbf{x}} = \mathbf{K}^{-1}\mathbf{x}, \bar{\mathbf{x}}' = \mathbf{K}^{-1}\mathbf{x}'$,

$$\mathbf{x}'^{\intercal}\mathbf{K}'^{-\intercal}\mathbf{E}\mathbf{K}^{-1}\mathbf{x} = 0 \tag{250}$$

this relationship holds between the Fundamental Matrix $\mathbf{F}$ and the Essential Matrix $\mathbf{E}$.

$$\begin{aligned} \mathbf{F} &= \mathbf{K}'^{-\intercal}\mathbf{E}\mathbf{K}^{-1} \\ \mathbf{E} &= \mathbf{K}'^{\intercal}\mathbf{F}\mathbf{K} \end{aligned} \tag{251}$$

Properties of the essential matrix

- **The Essential Matrix $\mathbf{E} = \mathbf{t}^{\wedge}\mathbf{R}$ has 5 degrees of freedom (DOF).** Both rotation $\mathbf{R}$ and translation $\mathbf{t}$ have 3 degrees of freedom each, but like the Fundamental Matrix, they lose 1 degree of freedom due to the homogeneous property, which introduces a scale ambiguity. This reduced degree of freedom forms additional constraints in the Essential Matrix compared to the Fundamental Matrix.
- **When the Essential Matrix is decomposed using Singular Value Decomposition (SVD), the diagonal matrix $\mathbf{D} = \begin{bmatrix} \sigma_1 & & \\ & \sigma_2 & \\ & & \sigma_3 \end{bmatrix}$ should have the two largest singular values σ_1, σ_2 equal, and the third singular value $\sigma_3 = 0$ should be satisfied.** More details are explained in the following section.

Result 9.17

The necessary and sufficient condition for any square matrix $\mathbf{E} \in \mathbb{R}^{3\times3}$ to be an Essential Matrix is that the two largest singular values σ_1, σ_2 are equal and the third singular value $\sigma_3 = 0$.

Proof

Forward Proof: Suppose there is a Fundamental Matrix $\mathbf{F}$ corresponding to the camera matrix pair $(\mathbf{P}, \mathbf{P}') = ([\mathbf{I}|0],\ [\mathbf{R}|\mathbf{t}])$

$$\mathbf{F} = \mathbf{t}^{\wedge}\mathbf{R} = \mathbf{S}\mathbf{R} \tag{252}$$

It follows that when the skew-symmetric matrix $\mathbf{t}^{\wedge} = \mathbf{S}$ has rank 2, $\mathbf{S}$ can always be transformed into the following form by changing the basis.

$$\mathbf{S} = k\mathbf{U}\mathbf{Z}\mathbf{U}^{\intercal} \tag{253}$$

Here, k represents an arbitrary scale value and is generally not considered. $\mathbf{U}$ is an arbitrary orthogonal matrix, and $\mathbf{Z}$ is a skew-symmetric matrix. To modify $\mathbf{U}\mathbf{Z}\mathbf{U}^{\intercal}$ to fit the form of SVD, a bit of algebraic trick is used. The skew-symmetric matrix $\mathbf{Z}$ and the orthogonal matrix $\mathbf{W}$ are defined as follows.

$$
\begin{aligned}
\mathbf{Z} &= \begin{bmatrix} & 1 & \\ -1 & & \\ & & \end{bmatrix} \\
\mathbf{W} &= \begin{bmatrix} & -1 & \\ 1 & & \\ & & 1 \end{bmatrix}
\end{aligned} \tag{254}
$$

There are the following useful properties between these matrices. These properties will frequently appear in the proof process of this section, so it is recommended to familiarize with them.

$$
\begin{aligned}
\mathbf{Z} &= \begin{bmatrix} 1 & & \\ & 1 & \\ & & \end{bmatrix} \mathbf{W} \quad \text{up to sign} \\
&= \text{diag}(1,1,0)\mathbf{W} \quad \text{up to sign}
\end{aligned} \tag{255}
$$

$$
\begin{aligned}
\mathbf{ZW} &= -\mathbf{ZW}^\intercal \\
&= \begin{bmatrix} 1 & & \\ & 1 & \\ & & \end{bmatrix} \\
&= \text{diag}(1,1,0)
\end{aligned} \tag{256}
$$

$$
\begin{aligned}
\mathbf{W}^\intercal &= \begin{bmatrix} -1 & & \\ & -1 & \\ & & 1 \end{bmatrix} \mathbf{W} \\
\mathbf{WW}^\intercal &= \mathbf{I}
\end{aligned} \tag{257}
$$

According to the above formula, $\mathbf{UZU}^\intercal = \mathbf{U}\text{diag}(1,1,0)\mathbf{WU}^\intercal$ can be represented as follows. Using this, the Fundamental Matrix $\mathbf{F}$ can be represented as follows.

$$
\begin{aligned}
\mathbf{F} &= \mathbf{SR} \\
&= \mathbf{UZU}^\intercal\mathbf{R} \\
&= \mathbf{U}\text{diag}(1,1,0)\mathbf{WU}^\intercal\mathbf{R} \\
&\sim \mathbf{U}\text{diag}(1,1,0)\mathbf{V}^\intercal \quad \text{up to similarity}
\end{aligned} \tag{258}
$$

According to the above equation, **the two largest singular values of F are the same, and the last singular value is zero. Therefore, the Fundamental Matrix F satisfies the properties of the Essential Matrix, thus becoming the Essential Matrix**.

Reverse Proof: Conversely, if the singular value decomposition (SVD) of an arbitrary square matrix $\mathbf{E} \in \mathbb{R}^{3\times 3}$ satisfies

$$
\mathbf{E} \sim \mathbf{U} \begin{bmatrix} 1 & & \\ & 1 & \\ & & 0 \end{bmatrix} \mathbf{V}^\intercal \tag{259}
$$

it becomes

$$
\begin{aligned}
\mathbf{E} &\sim \mathbf{U} \begin{bmatrix} 1 & & \\ & 1 & \\ & & 0 \end{bmatrix} \mathbf{V}^\intercal \\
&= \mathbf{U}\text{diag}(1,1,0)\mathbf{V}^\intercal \\
&= \mathbf{UZWV}^\intercal \qquad \because \mathbf{ZW} = \text{diag}(1,1,0) \\
&= \mathbf{UZU}^\intercal(\mathbf{UWV}^\intercal) \quad \because \mathbf{U}^\intercal\mathbf{U} = \mathbf{I} \\
&= \mathbf{SR} \\
&= \mathbf{t}^\wedge\mathbf{R}
\end{aligned} \tag{260}
$$

In the fourth row, the product of three orthogonal matrices $\mathbf{UWV}^\intercal$ meets the properties of a rotation matrix, therefore it can be denoted as $\mathbf{R}$. Thus, **at this time E becomes the Fundamental Matrix corresponding to the camera matrix pair $(\mathbf{P}, \mathbf{P}') = ([\mathbf{I}|0], [\mathbf{R}|\mathbf{t}])$.**

Extraction of cameras from the essential matrix

Result 9.18

When the Essential Matrix $\mathbf{E}$ is given as follows

$$\mathbf{E} = \mathbf{U}\begin{bmatrix}1 & & \\ & 1 & \\ & & 0\end{bmatrix}\mathbf{V}^{\intercal} \tag{261}$$

the matrix $\mathbf{E}$ can be decomposed into $\mathbf{R}, \mathbf{t}$ through SR Factorization. In this case, **$\mathbf{R}$ has two SR Factorization solutions with opposite signs (up to sign).**

$$\mathbf{R} = \mathbf{U}\mathbf{W}\mathbf{V}^{\intercal} \quad \text{or} \quad \mathbf{U}\mathbf{W}^{\intercal}\mathbf{V}^{\intercal} \tag{262}$$

Proof

As shown in (260), the Essential Matrix $\mathbf{E}$ can be decomposed into two matrices as follows.

$$\begin{aligned}\mathbf{E} &= (\mathbf{U}\mathbf{Z}\mathbf{U}^{\intercal})(\mathbf{U}\mathbf{W}\mathbf{V}^{\intercal}) = \mathbf{S}_0\mathbf{R}_0 \\ &= (\mathbf{U}\mathbf{Z}\mathbf{U}^{\intercal})(\mathbf{U}\mathbf{W}^{\intercal}\mathbf{V}^{\intercal}) = \mathbf{S}_0\mathbf{R}_0'\end{aligned} \tag{263}$$

In the second row, $(\mathbf{U}\mathbf{W}^{\intercal}\mathbf{V}^{\intercal})$ is the matrix obtained by expressing $\mathrm{diag}(1,1,0)$ using $-\mathbf{Z}\mathbf{W}^{\intercal}$ (up to sign). To prove that the Essential Matrix $\mathbf{E}$ is only decomposed into these two cases,

$$\mathbf{E} = \mathbf{S}_0\mathbf{R}_0 = \mathbf{S}\mathbf{R} \tag{264}$$

it is necessary to prove that $\mathbf{S} = \mathbf{S}_0, \mathbf{R} = \mathbf{R}_0 = \mathbf{R}_0'$.

The skew-symmetric matrices $\mathbf{S}, \mathbf{S}_0$ being rank 2 matrices can be represented in vector form as $\mathbf{S}_0^{\intercal} = \mathbf{s}_0^{\wedge}, \mathbf{S}^{\intercal} = \mathbf{s}^{\wedge}$

$$\mathbf{s}, \mathbf{s}_0 \in \mathrm{Nul}\ \mathbf{E}^{\intercal} \tag{265}$$

and thus $\mathbf{s}, \mathbf{s}_0$ are in a proportional relationship.

$$\mathbf{s} = \alpha\mathbf{s}_0 \quad \alpha \neq 0 \in \mathbb{R} \tag{266}$$

According to the above equation, $\mathbf{S}\mathbf{R} = \alpha\mathbf{S}_0\mathbf{R}_0$ holds and **since $\|\mathbf{R}\| = \|\mathbf{R}_0\| = 1$, by comparing the Frobenius Norm, it can be seen that $\alpha = \pm 1$.**

$$\mathbf{S} = \pm\mathbf{S}_0 \tag{267}$$

Next, the fact that $\mathbf{R} = \mathbf{R}_0 = \mathbf{R}_0'$ is proven. According to the results obtained so far, the Essential Matrix $\mathbf{E}$ can be written as follows.

$$\begin{aligned}\mathbf{E} &= \mathbf{U}\begin{bmatrix}1 & & \\ & 1 & \\ & & 0\end{bmatrix}\mathbf{V}^{\intercal} \\ &= \mathbf{S}_0\mathbf{R} \\ &= \mathbf{S}_0(\mathbf{U}\mathbf{X}\mathbf{V}^{\intercal}) \quad \because \mathbf{R} = \mathbf{U}\mathbf{X}\mathbf{V}^{\intercal} \\ &= \mathbf{U}\mathbf{Z}\mathbf{U}^{\intercal}\mathbf{U}\mathbf{X}\mathbf{V}^{\intercal} \quad \because \mathbf{U}^{\intercal}\mathbf{U} = \mathbf{I} \\ &= \mathbf{U}(\mathbf{Z}\mathbf{X})\mathbf{V}^{\intercal}\end{aligned} \tag{268}$$

Here, the rotation matrix $\mathbf{R}$ can be decomposed into three different orthogonal matrices as $\mathbf{R} = \mathbf{U}\mathbf{X}\mathbf{V}^{\intercal}$.

$\mathbf{Z}\mathbf{X} = \begin{bmatrix} & 1 & \\ -1 & & \\ & & \end{bmatrix}$ $\mathbf{X} = \mathrm{diag}(1,1,0)$ thus expanding this gives

$$\begin{aligned}\mathbf{X} &= \begin{bmatrix} & -1 & \\ 1 & & \\ & & \pm 1\end{bmatrix} \\ &= \mathbf{W} = -\mathbf{W}^{\intercal}\end{aligned} \tag{269}$$

Thus **therefore, including the change in sign (up to sign), $\mathbf{R} = \mathbf{R}_0 = \mathbf{R}_0'$ holds.**

$$\mathbf{R} = \mathbf{U}\mathbf{W}\mathbf{V}^\intercal \quad \text{or} \quad \mathbf{U}\mathbf{W}^\intercal\mathbf{V}^\intercal \tag{270}$$

Next, let's determine $\mathbf{t}$. Since the skew-symmetric matrix $\mathbf{S} = \mathbf{U}\text{diag}(1,1,0)\mathbf{U}^\intercal = \mathbf{t}^\wedge$, the following holds.

$$\mathbf{S}\mathbf{t} = \mathbf{t}^\wedge\mathbf{t} = 0 \tag{271}$$

In the above equation, **the solution for $\mathbf{t}$ becomes the Null Space of $\mathbf{S}$, thus it is the third column of matrix $\mathbf{U}$, denoted as $\mathbf{u}_3$.** However, since $\mathbf{S} = \pm\mathbf{S}_0$,

$$\mathbf{t} = \pm\mathbf{u}_3 \tag{272}$$

This determines the exact value of $\mathbf{t}$.

Result 9.19

Therefore, $\mathbf{t} = \pm\mathbf{u}_3$ and the previously determined $\mathbf{R} = \mathbf{U}\mathbf{W}\mathbf{V}^\intercal$ or $\mathbf{R} = \mathbf{U}\mathbf{W}^\intercal\mathbf{V}^\intercal$ give a total of four possible cases for the Essential Matrix $\mathbf{E} = \mathbf{t}^\wedge\mathbf{R}$.

Given two camera matrices $\mathbf{P}, \mathbf{P}'$ and the Essential Matrix $\mathbf{E}$, if $\mathbf{P} = [\mathbf{I} \mid 0]$ then the following four solutions exist for $\mathbf{P}'$.

$$\mathbf{P}' = [\mathbf{U}\mathbf{W}\mathbf{V}^\intercal| + \mathbf{u}_3] \text{ or } [\mathbf{U}\mathbf{W}\mathbf{V}^\intercal| - \mathbf{u}_3] \text{ or } [\mathbf{U}\mathbf{W}^\intercal\mathbf{V}^\intercal| + \mathbf{u}_3] \text{ or } [\mathbf{U}\mathbf{W}^\intercal\mathbf{V}^\intercal| - \mathbf{u}_3] \tag{273}$$

Geometrical interpretation of the four solutions

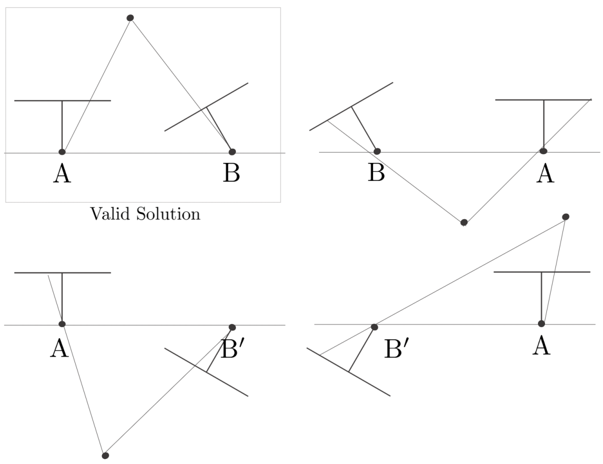
In the four solutions, the first two only differ by the sign of $\mathbf{u}_3$, indicating that the first and second cameras are flipped in opposite directions.

The relationship between the first and third solutions is as follows.

$$\begin{aligned}
\mathbf{E} &= [\mathbf{U}\mathbf{W}^\intercal\mathbf{V}^\intercal \mid \mathbf{u}_3] && \cdots \text{3rd solution} \\
&= [\mathbf{U}(\mathbf{W}\mathbf{V}^\intercal\mathbf{V}\mathbf{W}^\intercal)\mathbf{W}^\intercal\mathbf{V}^\intercal \mid \mathbf{u}_3] && \because \mathbf{W}\mathbf{V}^\intercal\mathbf{V}\mathbf{W}^\intercal = \mathbf{I} \\
&= [\mathbf{U}\mathbf{W}\mathbf{V}^\intercal \mid \mathbf{u}_3] \begin{bmatrix} \mathbf{V}\mathbf{W}^\intercal\mathbf{W}^\intercal\mathbf{V}^\intercal & \\ & 1 \end{bmatrix} \\
&= [\mathbf{U}\mathbf{W}\mathbf{V}^\intercal \mid \mathbf{u}_3] \begin{bmatrix} \mathbf{V}\begin{pmatrix} -1 & & \\ & -1 & \\ & & 1 \end{pmatrix}\mathbf{V}^\intercal & \\ & 1 \end{bmatrix} && \cdots \text{1st solution} \cdot [*]
\end{aligned} \tag{274}$$

Here $\mathbf{V}\mathbf{W}^\intercal\mathbf{W}^\intercal\mathbf{V}^\intercal$ becomes a matrix that vertically mirrors the two perpendicular directions of the Baseline. That is, **it forms a 180-degree rotation about the Baseline**.

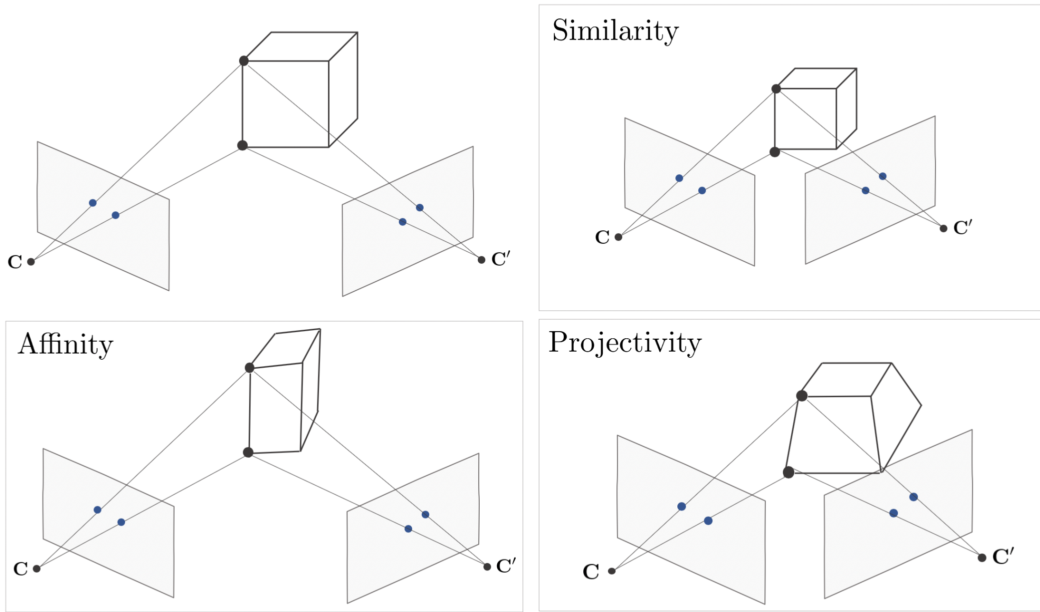
Four Solutions of Essential Matrix



The four solutions are geometrically represented as shown in the above figure. **Although mathematically a total of four solutions are derived, only one valid value exists in reality. Therefore,**

by selecting the unique solution where the 3D point $\mathbf{X}$ exists in front of both cameras, the $\mathbf{R}, \mathbf{t}$ decomposed from the Essential Matrix $\mathbf{E}$ can be successfully obtained.

7 3D Reconstruction of Cameras and Structure



As mentioned in the previous section, **the Fundamental Matrix $\mathbf{F}$ has projective ambiguity for various camera matrix pairs $(\mathbf{P}, \mathbf{P}')$, thus the 3D points in space calculated through the resulting $\mathbf{P}, \mathbf{P}'$ also have ambiguity.** This section explains how to remove ambiguity using Scene Constraint and Internal Constraint.

The projective reconstruction theorem

Theorem 10.1 (Projective reconstruction theorem)

Assuming sufficient pairs of corresponding points on the image planes of two cameras, $\mathbf{x}, \mathbf{x}'$, have been provided to compute the Fundamental Matrix $\mathbf{F}$, in cases where the following two sets of camera matrix pairs

$$\begin{aligned}(\mathbf{P}_1, \mathbf{P}'_1, \{\mathbf{X}_{1,i}\})\\(\mathbf{P}_2, \mathbf{P}'_2, \{\mathbf{X}_{2,i}\})\end{aligned} \tag{275}$$

share the same Fundamental Matrix $\mathbf{F}$, there necessarily exists a Homography matrix $\mathbf{H} \in \text{PGL}_4$ that satisfies

$$(\mathbf{P}_2, \mathbf{P}'_2, \{\mathbf{X}_{2,i}\}) = \mathbf{H} \cdot (\mathbf{P}_1, \mathbf{P}'_1, \{\mathbf{X}_{1,i}\}) \tag{276}$$

Here, the $\cdot$ operation is defined as follows.

$$\begin{aligned}\mathbf{P}_2 &= \mathbf{H} \cdot \mathbf{P}_1 = \mathbf{P}_1\mathbf{H}^{-1}\\ \mathbf{P}'_2 &= \mathbf{H} \cdot \mathbf{P}'_1 = \mathbf{P}'_1\mathbf{H}^{-1}\\ \mathbf{X}_{2i} &= \mathbf{H} \cdot \mathbf{X}_{1i} = \mathbf{H}\mathbf{X}_{1i}\end{aligned} \tag{277}$$

Proof

In the previous section, we proved that there exists a homography matrix $\mathbf{H} \in \mathbb{R}^{4\times4}$ such that $\mathbf{P}_2 = \mathbf{P}_1\mathbf{H}^{-1}$ and $\mathbf{P}'_2 = \mathbf{P}'_1\mathbf{H}^{-1}$ between pairs of camera matrices sharing the same Fundamental Matrix $\mathbf{F}$. Applying this results in

$$
\begin{aligned}
\mathbf{P}_2\mathbf{X}_{2i} &= \mathbf{P}_1\mathbf{H}^{-1}\mathbf{X}_{2i} = \mathbf{x}_i = \mathbf{P}_1\mathbf{X}_{1i} \\
\mathbf{P}'_2\mathbf{X}_{2i} &= \mathbf{P}'_1\mathbf{H}^{-1}\mathbf{X}_{2i} = \mathbf{x}'_i = \mathbf{P}'_1\mathbf{X}_{1i}
\end{aligned} \tag{278}
$$

If we define the back-projection lines from $\mathbf{P}_1, \mathbf{P}_2$ as $\mathcal{R}$ and those from $\mathbf{P}'_1, \mathbf{P}'_2$ as $\mathcal{R}'$, then

$$
\{\mathbf{H}^{-1}\mathbf{X}_{2i}, \mathbf{X}_{1i}\} \in \mathcal{R} \cap \mathcal{R}' \tag{279}
$$

The points in 3D space $\{\mathbf{H}^{-1}\mathbf{X}_{2i}, \mathbf{X}_{1i}\}$ are the intersection of these lines. Therefore, excluding the case where the lines $\mathcal{R}, \mathcal{R}'$ are the same as the baseline, $\{\mathbf{H}^{-1}\mathbf{X}_{2i}, \mathbf{X}_{1i}\}$ represents a single intersection point.

$$
\begin{aligned}
&\mathbf{H}^{-1}\mathbf{X}_{2i} = \mathbf{X}_{1i} \\
&\mathbf{X}_{2i} = \mathbf{H}\mathbf{X}_{1i}
\end{aligned} \tag{280}
$$

Thus, without additional geometric principles, in such scenarios, it is possible to restore up to projectivity.

Stratified reconstruction

To address the projective ambiguity problem of the camera matrix pairs $(\mathbf{P}, \mathbf{P}')$ computed using the Fundamental Matrix $\mathbf{F}$, assume that there exists a ground truth pair of camera matrices

$$
(\mathbf{P}_0, \mathbf{P}'_0, \{\hat{\mathbf{X}}_i\}) \tag{281}
$$

Then, there exists a homography matrix $\mathbf{H}$ satisfying

$$
\begin{aligned}
&(\mathbf{P}_1, \mathbf{P}'_1, \{\mathbf{X}_i\}) = \mathbf{H} \cdot (\mathbf{P}_0, \mathbf{P}'_0, \{\hat{\mathbf{X}}_i\}) \\
&\mathcal{T} = \mathbf{H} \cdot \mathcal{T}_0
\end{aligned} \tag{282}
$$

During the affine reconstruction phase, a homography matrix $\mathbf{H}_A$ that preserves the characteristics of parallel lines is identified.

$$
\mathbf{H}_A(\pi_\infty) = \pi_\infty \tag{283}
$$

After finding the vanishing point using the set of points $\{\mathbf{X}_i\}$ in 3D space, the projected plane of infinity π'_∞ in the image plane can be calculated. If π'_∞ is represented as

$$
\pi'_\infty = \begin{bmatrix} a & b & c & 1 \end{bmatrix}^\intercal \tag{284}
$$

The homography matrix $\mathbf{H}_A$ that sends the actual plane at infinity $\pi_\infty = \begin{pmatrix} 0 & 0 & 0 & 1 \end{pmatrix}^\intercal$ to π'_∞ can be formulated as

$$
\mathbf{H}_A = \begin{bmatrix} \mathbf{I}_3 \mid 0 \\ \pi'^\intercal_\infty \end{bmatrix} = \begin{bmatrix} & \mathbf{I}_3 & & 0 \\ a & b & c & 1 \end{bmatrix} \tag{285}
$$

In this case, the relationship $\mathbf{H}_A^\intercal \pi_\infty = \pi'_\infty$ holds true.

$$
\pi_\infty = \mathbf{H}_A^{-\intercal} \pi'_\infty \tag{286}
$$

Applying this to $\mathcal{T}$ results in

$$
\begin{aligned}
&\mathcal{T}_a = \mathbf{H} \cdot \mathcal{T} \\
&(\mathbf{P}\mathbf{H}_A^{-1}, \mathbf{P}'\mathbf{H}_A^{-1}, \{\mathbf{H}_A\mathbf{X}_i\}) = \mathbf{H} \cdot (\mathbf{P}, \mathbf{P}', \{\mathbf{X}_i\})
\end{aligned} \tag{287}
$$

This operation allows for affine reconstruction of the image. At this stage, the ground truth $\mathcal{T}_0$ and $\mathcal{T}_a$ differ only by an affine transformation, meaning all parallel lines have been restored, but orthogonal lines have not yet been reconstructed.

The step to metric reconstruction

In the metric reconstruction step, we must find the homography matrix $\mathbf{H}_S$ that preserves orthogonal lines.

$$\mathbf{H}_S(\Omega_\infty) = \Omega_\infty \tag{288}$$

Here, Ω_∞ represents the Absolute Conic existing on the plane at infinity. In the affine reconstruction step, $\mathcal{T}_a$ positions the infinite plane π_∞ at actual infinity. Using orthogonal points among $\{\mathbf{X}_i\}$, we can find Ω_∞. If the corresponding pairs of orthogonal points are $\mathbf{d}_1, \mathbf{d}_2$, then

$$\mathbf{d}_1^\intercal \omega \mathbf{d}_2 = 0 \tag{289}$$

satisfies, and ω can be found as the image of the Absolute Conic Ω_∞ (Image of Absolute Conic, IAC) projected onto the image plane.

When the camera matrix $\mathbf{P} = \mathbf{K}[\mathbf{R}|\mathbf{t}] = [\mathbf{M} \mid \mathbf{m}]$ is given, the infinite plane π_∞ is projected onto the image plane π by $\mathbf{M}$.

$$\mathbf{M} : \pi_\infty \mapsto \pi \tag{290}$$

Conversely, $\mathbf{M}^{-1}$ maps points on the image plane π to the infinite plane π_∞.

$$\mathbf{M}^{-1} : \pi \mapsto \pi_\infty \tag{291}$$

Applying $\mathbf{M}^{-1}$ to the Absolute Conic projected onto the image plane results in $\mathbf{M}^{-1}(\omega) = \tilde{\Omega}_\infty$. Since $\tilde{\Omega}_\infty$ is not yet metrically reconstructed, $\Omega_\infty \neq \tilde{\Omega}_\infty$. The following formula holds between them.

$$\mathbf{H}_S(\tilde{\Omega}_\infty) = \Omega_\infty = \mathbf{I}_3 \tag{292}$$

Result 10.5

Calling the Absolute Conic projected onto the image plane ω and the camera matrix obtained from affine reconstruction $\mathbf{P} = [\mathbf{M} \mid \mathbf{m}]$, the Homography $\mathbf{H}_S$ performing metric reconstruction can be calculated as follows.

$$\mathbf{H}_S = \begin{bmatrix} \mathbf{A}^{-1} & \\ & 1 \end{bmatrix} \tag{293}$$

Here, $\mathbf{A}\mathbf{A}^\intercal = (\mathbf{M}^\intercal \omega \mathbf{M})^{-1}$, and $\mathbf{A}$ can be obtained through Cholesky Decomposition.

Proof

Calling the Homography performing metric reconstruction $\mathbf{H}_S = \begin{bmatrix} \mathbf{A}^{-1} & \\ & 1 \end{bmatrix}$, and assuming the image plane to be π_s and the camera matrix $\mathbf{P}_s$, the relationship $\mathbf{H}_S(\pi) = \pi_s$ holds, and the formula transforming the infinite plane by Homography is

$$\begin{aligned} \pi_\infty \mathbf{A}^{-1} &= \pi_{\infty,s} \\ \mathbf{H}_S|_{\pi_\infty} &= \mathbf{A}^{-1} \end{aligned} \tag{294}$$

Transformations of the infinite planes $\pi_\infty, \pi_{\infty,s}$ onto the image planes π, π_s are as follows.

$$\begin{aligned} \mathbf{P}|_{\pi_\infty} &= \mathbf{M} \\ \mathbf{P}_s|_{\pi_\infty} &= \mathbf{M}\mathbf{A} \end{aligned} \tag{295}$$

Thus, $\mathbf{H}_S(\tilde{\Omega}_\infty)$ is as follows.

$$\begin{aligned} \mathbf{H}_S(\tilde{\Omega}_\infty) &= (\mathbf{A}^{-1})^{-\intercal} \tilde{\Omega}_\infty \mathbf{A} \\ &= \mathbf{A}^\intercal (\mathbf{M}^\intercal \omega \mathbf{M}) \mathbf{A} \end{aligned} \tag{296}$$

Thus, by the property of the Absolute Conic, $\mathbf{A}^\intercal(\mathbf{M}^\intercal \omega \mathbf{M})\mathbf{A}$ must equal $\mathbf{I}$, so

$$\begin{aligned} \mathbf{A}^\intercal(\mathbf{M}^\intercal \omega \mathbf{M})\mathbf{A} &= \mathbf{I} \\ (\mathbf{M}^\intercal \omega \mathbf{M})^{-1} &= \mathbf{A}\mathbf{A}^\intercal \end{aligned} \tag{297}$$

Therefore, to construct $\mathbf{H}_S = \begin{bmatrix} \mathbf{A}^{-1} & \\ & 1 \end{bmatrix}$, we must use the properties of orthogonal lines on the image plane to find ω and then obtain $\mathbf{A}\mathbf{A}^\intercal$ through Cholesky Decomposition of $(\mathbf{M}^\intercal \omega \mathbf{M})^{-1}$. This allows us to determine the value of $\mathbf{A}^{-1}$ and ultimately construct the Homography $\mathbf{H}_S$ that performs Metric Reconstruction.

8 Computation of the Fundamental Matrix F

Basic equations

When there are corresponding point pairs $\mathbf{x} \leftrightarrow \mathbf{x}'$ between two cameras,

$$\mathbf{x}'^\intercal \mathbf{F} \mathbf{x} = 0 \tag{298}$$

the 3x3 matrix $\mathbf{F}$ that satisfies this is called the Fundamental Matrix. If three pairs of corresponding points $\mathbf{x} = (x_1, x_2, x_3)$, $\mathbf{x}' = (x'_1, x'_2, x'_3)$ are given, $\mathbf{x}'^\intercal \mathbf{F} \mathbf{x} = 0$ can be rearranged to $\sum \mathbf{x}'_i \mathbf{x}_j \mathbf{f}_{ij} = 0$. This equation can be expressed as a linear system

$$\mathbf{A}\mathbf{f} = 0 \tag{299}$$

where $\mathbf{A}$ and $\mathbf{f}$ are expanded as follows:

$$\begin{aligned} &\mathbf{A} = (\mathbf{a}_{ik}) \in \mathbb{R}^{N \times 9}, \quad \text{i-th row is } (x'_{i1}x_{i1}, x'_{i1}x_{i2}, \cdots, x'_{i3}x_{i3}) \\ &\mathbf{f} = (\mathbf{f}_{11}, \mathbf{f}_{12}, \mathbf{f}_{13}, \cdots, \mathbf{f}_{33})^\intercal \in \mathbb{R}^{9 \times 1} \end{aligned} \tag{300}$$

If the number of given corresponding point pairs N is greater than 8 and the data is generated from perfect, noise-free data, then $\mathbf{A}$ becomes a rank 8 matrix and a unique solution $\mathbf{f}$ exists. At this point, $\mathbf{f}$ becomes an element of the Null Space of $\mathbf{A}$.

However, as data always contains noise in most cases, the matrix $\mathbf{A}$ becomes a full rank matrix of rank 9. Then, if a solution is found, $\mathbf{f}$ always computes to the zero vector, so we must find an approximate solution $\mathbf{f}$ that minimizes the magnitude of $\|\mathbf{A}\mathbf{f}\|$ under the condition $\|\mathbf{f}\| = 1$. The method to find the solution is to decompose $\mathbf{A}$ using Singular Value Decomposition (SVD) and then

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^\intercal \tag{301}$$

consider the column vector of $\mathbf{V}$ corresponding to the smallest absolute value of the eigenvalue in the diagonal matrix $\mathbf{D}$ as the general solution for $\mathbf{f}$. This column vector $\mathbf{f}$ minimizes the magnitude of $\|\mathbf{A}\mathbf{f}\|$ and is referred to as the Linear Solution.

However, **the matrix $\mathbf{F}_0$ restored by the Linear Solution f is not guaranteed to be a rank 2 matrix.** Numerically derived solutions typically yield a rank 3 matrix. Thus, another SVD must be used to find the closest rank 2 matrix to $\mathbf{F}_0$.

$$\mathbf{F}_0 = \mathbf{U}_0 \mathbf{D}_0 \mathbf{V}_0^\intercal \tag{302}$$

At this point, the diagonal matrix $\mathbf{D}_0 = \begin{bmatrix} \sigma_1 & & \\ & \sigma_2 & \\ & & \sigma_3 \end{bmatrix}$ typically has the last singular value σ_3 not equal to zero, thus making it a rank 3 matrix. Therefore, **arbitrarily setting $\sigma_3 = 0$ results in the closest rank 2 matrix to $\mathbf{F}_0$.**

The minimum case - seven point correspondences

In cases where seven corresponding point pairs are given, $\mathbf{A} \in \mathbb{R}^{7 \times 9}$ becomes a rank 7 matrix. Thus, the dimension of the Null Space is 2, resulting in infinitely many solutions. From $\mathbf{A}\mathbf{f} = 0$, two linearly independent Fundamental Matrices $\mathbf{F}_1, \mathbf{F}_2$ are derived and

$$\mathbf{F} = \alpha \mathbf{F}_1 + (1 - \alpha)\mathbf{F}_2 \tag{303}$$

is used to compute $\mathbf{F}$. Since the rank of the matrix $\mathbf{F}$ is 2, $\det(\mathbf{F}) = 0$ must be satisfied, leading to a cubic equation for α. Cases with three real solutions for the cubic equation become Degenerate Configurations.

The normalized 8-point algorithm

The matrix $\mathbf{A}$ is composed of the given corresponding point pairs $\mathbf{x} \leftrightarrow \mathbf{x}'$, therefore **if the solution is computed directly without normalizing the corresponding point pairs, such as $\mathbf{x}_i = (1000000, 20000000, 1)$ where the first two values are significantly larger than the last, it results in a numerically unstable problem**. Hence, Homography transformations $\mathbf{H}, \mathbf{H}'$ that set the Centroid to 0 and the average distance from the Centroid to $\sqrt{2}$ are applied to $\mathbf{x}_i, \mathbf{x}'_i$.

$$\begin{aligned} \mathbf{x}_i &\rightarrow \mathbf{H}\mathbf{x}_i \\ \mathbf{x}'_i &\rightarrow \mathbf{H}'\mathbf{x}'_i \end{aligned} \tag{304}$$

Using these transformed corresponding point pairs, a numerically stable Fundamental Matrix $\mathbf{F}'$ can be calculated. Following the described content, $\mathbf{F}'$ is computed and then restored to the original corresponding point pairs $\mathbf{x} \leftrightarrow \mathbf{x}'$ by

$$\begin{aligned} 0 &= (\mathbf{H}'\mathbf{x}')^{\intercal}\mathbf{F}'(\mathbf{H}\mathbf{x}) \\ &= \mathbf{x}^{\intercal}(\mathbf{H}'^{\intercal}\mathbf{F}'\mathbf{H})\mathbf{x} \end{aligned} \tag{305}$$

thus, $\mathbf{F} = \mathbf{H}'^{\intercal}\mathbf{F}'\mathbf{H}$ can be computed as the Fundamental Matrix for the original corresponding point pairs.

Degenrate configurations

Two camera matrix correspondences $(\mathbf{P}, \mathbf{P}', \{\mathbf{X}_i\})$, $(\mathbf{Q}, \mathbf{Q}', \{\mathbf{Y}_i\})$ are considered to be in **Conjugate Configuration** if they satisfy the following condition:

$$\begin{aligned} &\mathbf{P}\mathbf{X}_i = \mathbf{Q}\mathbf{Y}_i, \quad \mathbf{P}'\mathbf{x}_i = \mathbf{Q}'\mathbf{Y}' \quad {}^{\forall}s \\ &(\mathbf{P}, \mathbf{P}), (\mathbf{Q}, \mathbf{Q}') \;\; \text{share the same} \;\; \mathbf{F} \end{aligned} \tag{306}$$

In this case, **camera matrix correspondences $(\mathbf{P}, \mathbf{P}', \{\mathbf{X}_i\})$ that have a Conjuate Configuration are called a Critical Triple.**

The necessary and sufficient condition for an arbitrary camera correspondence $\mathbf{P}, \mathbf{P}', \{\mathbf{X}_i\}$ to be a Critical Triple is that the camera centers $\mathbf{C}, \mathbf{C}'$ and the points in 3D space $\mathbf{X}_i$ must be included on a Ruled Quadric Surface, i.e., a Quadric Surface that includes a line.

Proof

To demonstrate that an arbitrary camera matrix correspondence $(\mathbf{P}, \mathbf{P}', \{\mathbf{X}_i\})$ is a Critical Triple, it is necessary to use the fact that the corresponding camera matrix pair $(\mathbf{Q}, \mathbf{Q}', \{\mathbf{Y}_i\})$ shares the same Fundamental Matrix $\mathbf{F}$. In $\mathbb{P}^3$ space, a quadric surface is defined by a 4x4 matrix.

$$\mathbf{S}_p := \mathbf{P}'\mathbf{F}_{QQ'}\mathbf{P} \tag{307}$$

Here, the quadric surface $\mathbf{S}_p \in \mathbb{R}^{4\times 4}$ must include the camera centers $\mathbf{C}, \mathbf{C}'$ and the points in 3D space $\{\mathbf{X}_i\}$.

$\mathbf{S}_p$ is a Ruled Quadric Surface, so it is sufficient to show that it includes the baseline, a line connecting the camera centers. It is demonstrated by showing that an arbitrary point $\mathbf{X}$ on the baseline satisfies

$$\mathbf{S}_p\mathbf{X} = \mathbf{P}'\mathbf{F}_{QQ'}\mathbf{P}\mathbf{X} = 0 \quad \because \mathbf{F}_{QQ'}\mathbf{P}\mathbf{X} = 0 \tag{308}$$

thus confirming that $\mathbf{S}_p$ is a Ruled Quadric Surface.

Conversely, if the camera centers $\mathbf{C}, \mathbf{C}'$ and the points in 3D space $\{\mathbf{X}_i\}$ belong to $\mathbf{S}_p$, then the following formula holds:

$$(\mathbf{P}'\mathbf{X}_i)^{\intercal}\mathbf{F}_{QQ'}(\mathbf{P}\mathbf{X}_i) = 0 \tag{309}$$

Here $\mathbf{x}' = \mathbf{P}'\mathbf{X}_i, \mathbf{x} = \mathbf{P}\mathbf{X}_i$, and since $\mathbf{F}_{QQ'}$, this means there exists a corresponding point pair $\mathbf{x}, \mathbf{x}'$. Therefore,

$$\begin{aligned} \mathbf{x}' &= \mathbf{P}'\mathbf{X}_i = \mathbf{Q}'\mathbf{Y}_i \\ \mathbf{x} &= \mathbf{P}\mathbf{X}_i = \mathbf{Q}\mathbf{Y}_i \end{aligned} \tag{310}$$

is satisfied, thus camera matrix correspondences $(\mathbf{P}, \mathbf{P}', \{\mathbf{X}_i\})$ and $(\mathbf{Q}, \mathbf{Q}', \{\mathbf{Y}_i\})$ share the same Fundamental Matrix $\mathbf{F} = \mathbf{F}_{PP'} = \mathbf{F}_{QQ'}$, confirming that they are a Conjugate Triple.

The Gold Standard method

Image plane corresponding point pairs $\mathbf{x} \leftrightarrow \mathbf{x}'$ generally contain noise, so it is not possible to accurately compute the Fundamental Matrix $\mathbf{F}$ and an approximate solution $\mathbf{F}'$ must be computed by minimizing the magnitude of $\|\mathbf{Af}\|$. Therefore, **to more accurately compute $\mathbf{F}'$, a method called The Gold Standard method is used to correct the given data x, x' to be closer to the Ground Truth $\hat{\mathrm{x}}, \hat{\mathrm{x}}'$, similar to the 2D Homography problem.**

- When more than eight corresponding point pairs $\mathbf{x} \leftrightarrow \mathbf{x}'$ are given, first, as previously described, a rank 2 Fundamental Matrix $\mathbf{F}_0$ is computed using two rounds of SVD. This is referred to as the Linear Solution.
- Next, the following formula holds for the Epipole $\mathbf{e}, \mathbf{e}'$:

$$\mathbf{e}'\mathbf{F}_0 = 0, \quad \mathbf{F}_0\mathbf{e} = 0 \tag{311}$$

- From the above formula, $\mathbf{e}, \mathbf{e}'$ are calculated.
- Once the Epipole is calculated, the Canonical Form of the camera matrices $\mathbf{P}, \mathbf{P}'$ can be computed.

$$\begin{aligned} &\mathbf{P} = [\mathbf{I}|0], \quad \mathbf{P}' = [\mathbf{M}|\mathbf{m}] \\ &\text{where,} \;\; \mathbf{M} = \mathbf{e}'^{\wedge}\mathbf{F}, \;\; \mathbf{m} = \mathbf{e}' \end{aligned} \tag{312}$$

- This allows for the calculation of Back-projection lines for $\mathbf{x}, \mathbf{x}'$.
- The closest point in 3D space $\mathbf{X}_i$ to these Back-projection lines can be computed.
- The 3D point $\mathbf{X}_i$ is then reprojected onto the image plane.

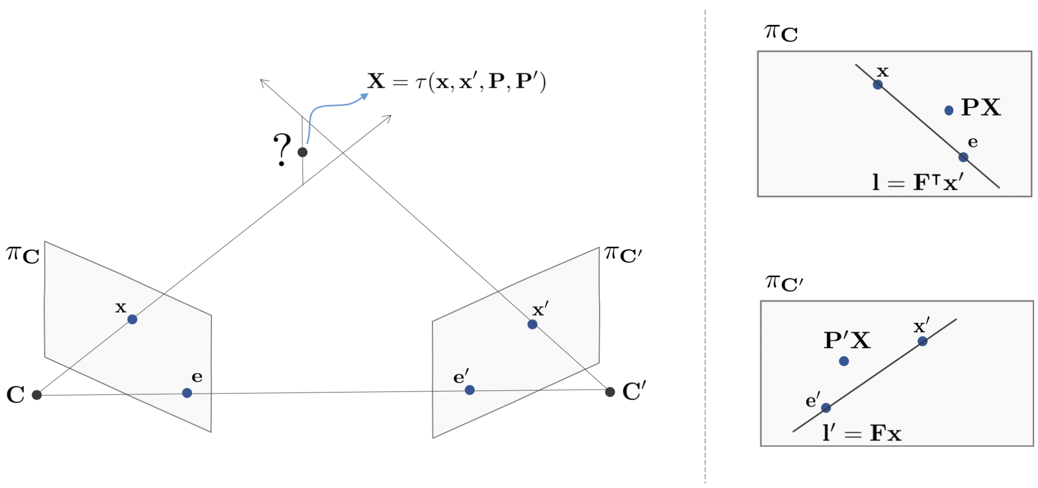
$$\begin{aligned} \hat{\mathrm{x}} &= \mathbf{P}\mathbf{X}_i \\ \hat{\mathrm{x}}' &= \mathbf{P}'\mathbf{X}_i \end{aligned} \tag{313}$$

The following is used to minimize the reprojection error using nonlinear optimization methods such as Gauss-Newton, Levenerg-Marquardt, etc.

$$\sum_i d(\mathbf{x}, \hat{\mathbf{x}})^2 + d(\mathbf{x}', \hat{\mathbf{x}}')^2 \tag{314}$$

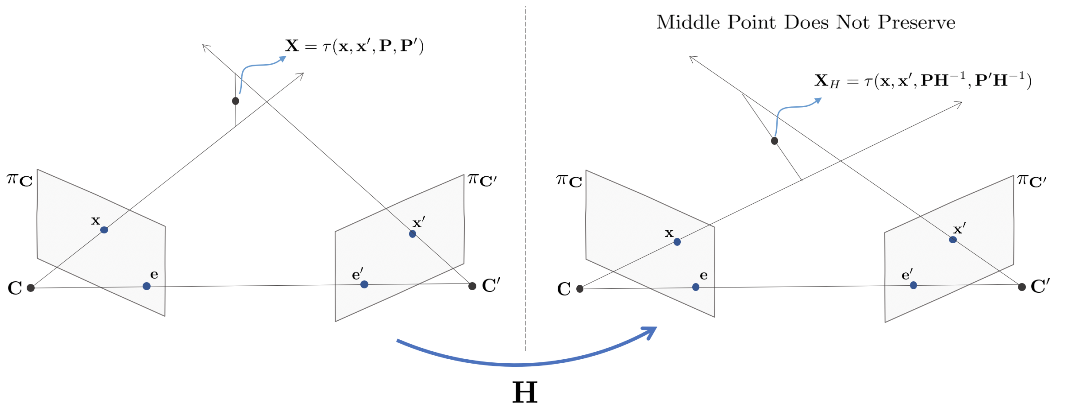
In this case, the parameters optimized are $\mathbf{X}_i$ and the camera matrix $\mathbf{P}' = [\mathbf{M}|\mathbf{m}]$.

9 Structure Computation



Problem statement

This section explains the method to calculate the point $\mathbf{X}$ in $\mathbb{P}^3$ space given a pair of camera matrices $(\mathbf{P}, \mathbf{P}')$ and corresponding image plane point pairs $(\mathbf{x}, \mathbf{x}')$. Theoretically, $\mathbf{X}$ can be computed through the intersection of two lines obtained by back-projecting the corresponding point pairs $\mathbf{x}, \mathbf{x}'$, but in reality, due to noise, the two back-projected lines do not intersect. Furthermore, noise also means that the Epipolar Lines $\mathbf{l}, \mathbf{l}'$ do not meet $\mathbf{x}, \mathbf{x}'$ on the image planes.



To solve this, a method exists that calculates $\mathbf{X}$ by determining the midpoint of the shortest line perpendicular to both back-projected lines. This method is suitable when $\mathbf{P}, \mathbf{P}'$ are fixed, such as in calibration methods using a checkerboard; however, in most cases, $\mathbf{P}, \mathbf{P}'$ are not uniquely determined due to projective ambiguity.

In the real world, camera matrices are known up to affinity or projectivity in most cases. For accurate computation of $\mathbf{X}$, triangulation must be invariant to projective ambiguity. Triangulation τ refers to the method of computing $\mathbf{X} \in \mathbb{P}^3$ from the given corresponding point pairs and camera matrix pairs.

$$\mathbf{X} = \tau(\mathbf{x}, \mathbf{x}', \mathbf{P}, \mathbf{P}') \tag{315}$$

For τ to be invariant to projective ambiguity, it must satisfy the following with respect to any Homography $\mathbf{H} \in \mathrm{PGL}_4$:

$$\begin{aligned} \mathbf{X}_H &= \tau(\mathbf{x}, \mathbf{x}', \mathbf{P}\mathbf{H}^{-1}, \mathbf{P}'\mathbf{H}^{-1}) \\ &= \mathbf{H} \cdot \tau(\mathbf{x}, \mathbf{x}', \mathbf{P}, \mathbf{P}') \\ &= \mathbf{H}\mathbf{X} \end{aligned} \tag{316}$$

In such invariant cases,

$$\begin{aligned} \mathbf{P}\mathbf{H}^{-1}\mathbf{X}_H &= \mathbf{P}\mathbf{H}^{-1}\mathbf{H}\mathbf{X} \\ &= \mathbf{P}\mathbf{X} \\ &= \mathbf{x} \end{aligned} \tag{317}$$

Similarly, $\mathbf{P}'\mathbf{H}^{-1}\mathbf{X}_H = \mathbf{x}'$. For example, when the Fundamental Matrix $\mathbf{F}$ and camera matrix pairs $(\mathbf{P}, \mathbf{P}')$, along with the corresponding point pairs $(\mathbf{x}, \mathbf{x}')$ are given, the camera matrix pairs transformed by Homography $(\mathbf{P}\mathbf{H}^{-1}, \mathbf{P}'\mathbf{H}^{-1})$ also share the same $\mathbf{F}$. Through this,

$$\begin{aligned} &\{\mathbf{X}\}: \quad \text{3D Points from } (\mathbf{x}, \mathbf{x}') \text{ and } (\mathbf{P}, \mathbf{P}'). \\ &\{\mathbf{X}_H\}: \quad \text{3D Points from } (\mathbf{x}, \mathbf{x}') \text{ and } (\mathbf{P}\mathbf{H}^{-1}, \mathbf{P}'\mathbf{H}^{-1}). \end{aligned} \tag{318}$$

From which two PointClouds $\{\mathbf{X}\}, \{\mathbf{X}_H\}$ can be derived, and for the PointCloud to be invariant under Homography, the transformation property

$$\mathbf{H}\{\mathbf{X}\} = \{\mathbf{X}_H\} \tag{319}$$

Must be satisfied. The aforementioned mid-point method becomes invalid under Homography transformation as the shortest distance line perpendicular to the two back-projected lines no longer remains orthogonal, and the midpoint does not stay central.

Linear triangulation methods

This section describes methods using linear equations to determine $\mathbf{X} \in \mathbb{P}^3$. From $\mathbf{x} = \mathbf{P}\mathbf{X}, \mathbf{x}' = \mathbf{P}'\mathbf{X}$,

$$
\begin{aligned}
\mathbf{x}^{\wedge}(\mathbf{P}\mathbf{X}) = 0 \\
\mathbf{x}'^{\wedge}(\mathbf{P}'\mathbf{X}) = 0
\end{aligned} \tag{320}
$$

The equation holds.

$$
\begin{aligned}
\mathbf{x} &= \begin{bmatrix} x & y & 1 \end{bmatrix}^{\intercal} \\
\mathbf{P} &= \begin{bmatrix} \mathbf{p}_{1,row}^{\intercal} & \mathbf{p}_{2,row}^{\intercal} & \mathbf{p}_{3,row}^{\intercal} \end{bmatrix}^{\intercal} \\
\mathbf{X} &= \begin{bmatrix} X & Y & Z & W \end{bmatrix}^{\intercal}
\end{aligned} \tag{321}
$$

When $\mathbf{p}_{i,row}^{\intercal}$ is the i-th row of $\mathbf{P}$, expanding $\mathbf{x}^{\wedge}(\mathbf{P}\mathbf{X})$ and simplifying yields

$$
\begin{aligned}
x(\mathbf{p}_{3,row}^{\intercal}\mathbf{X}) - (\mathbf{p}_{1,row}^{\intercal}\mathbf{X}) = 0 \\
y(\mathbf{p}_{3,row}^{\intercal}\mathbf{X}) - (\mathbf{p}_{2,row}^{\intercal}\mathbf{X}) = 0 \\
x(\mathbf{p}_{2,row}^{\intercal}\mathbf{X}) - y(\mathbf{p}_{1,row}^{\intercal}\mathbf{X}) = 0
\end{aligned} \tag{322}
$$

This is done along with constructing $\mathbf{x}'^{\wedge}(\mathbf{P}'\mathbf{X})$, and the system of linear equations for $\mathbf{X}$ is reformulated as

$$
\underbrace{\begin{bmatrix} x\mathbf{p}_{3,row}^{\intercal} - \mathbf{p}_{1,row}^{\intercal} \\ y\mathbf{p}_{3,row}^{\intercal} - \mathbf{p}_{2,row}^{\intercal} \\ x'\mathbf{p}_{3,row}'^{\intercal} - \mathbf{p}_{1,row}'^{\intercal} \\ y'\mathbf{p}_{3,row}'^{\intercal} - \mathbf{p}_{2,row}'^{\intercal} \end{bmatrix}}_{\mathbf{A}} \mathbf{X} = 0 \tag{323}
$$

Resulting in the form $\mathbf{A}\mathbf{X} = 0$. In the absence of noise, the rank of $\mathbf{A} \in \mathbb{R}^{4\times4}$ would be 3, allowing the unique $\mathbf{X}$ to be calculated through the Null Space vector. However, with noise present, the rank of $\mathbf{A}$ becomes 4, leading to infinitely many solutions. Therefore, $\mathbf{A}$ is decomposed using Singular Value Decomposition (SVD), and the approximate solution $\hat{\mathbf{X}}$ is calculated when $\|\mathbf{X}\| = 1$ to minimize $\|\mathbf{A}\mathbf{X}\|$.

Applying a Homography to the linear method, we have

$$
\begin{aligned}
y(\mathbf{p}_{3,row}) - \mathbf{p}_{2,row} \rightarrow & \, y(\mathbf{p}\mathbf{H}^{-1})_{3,row} - (\mathbf{p}\mathbf{H}^{-1})_{2,row} \\
&= y(\mathbf{p}_{3,row})\mathbf{H}^{-1} - \mathbf{p}_{2,row}\mathbf{H}^{-1} \\
&= (y(\mathbf{p}_{3,row}) - \mathbf{p}_{2,row})\mathbf{H}^{-1}
\end{aligned} \tag{324}
$$

Thus, $\mathbf{A} \Rightarrow \mathbf{A}\mathbf{H}^{-1}$, transforming $\mathbf{X} \Rightarrow \mathbf{X}_H$. Through this transformation,

$$
\|\mathbf{A}\mathbf{X}\| = \|\mathbf{A}\mathbf{H}^{-1}\mathbf{X}_H\| \tag{325}
$$

holds true. However, when performing the Homography transformation, $\|\mathbf{X}_H\| \neq 1$, failing to preserve the property $\|\mathbf{X}\| = 1$. Therefore, the method using linear equations is not invariant to Homography, making it not the optimal solution.

Inhomogeneous method

When the camera matrix pair $\mathbf{P}, \mathbf{P}'$ is determined up to affinity, we set $\mathbf{X} = (X, Y, Z, 1)^{\intercal}$ and find an approximate solution $\hat{\mathbf{X}}$ that minimizes $\|\mathbf{A}\mathbf{X}\|$. Since $\mathbf{X}$ is an Affine Point, the constraint $\|\mathbf{X}\| = 1$ disappears. Therefore, the **Inhomogeneous method** has the property of being invariant to any arbitrary Affine transformation $\mathbf{H}_A$.

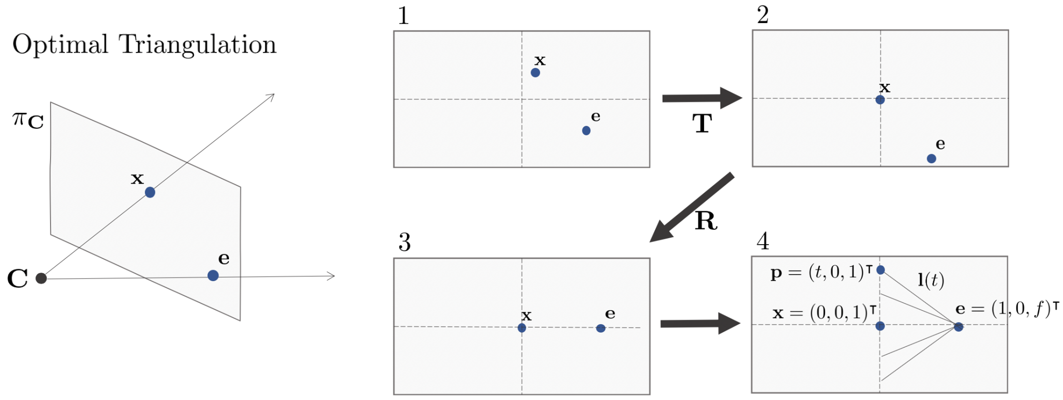
An optimal solution

This section explains the Optimal Triangulation method for finding the best $\mathbf{X} \in \mathbb{P}^3$.

- For a given camera matrix pair $(\mathbf{P}, \mathbf{P}')$ and corresponding point pair $(\mathbf{x}, \mathbf{x}')$, the approximate solution $\hat{\mathbf{X}}$ obtained through the linear equation $\mathbf{A}\mathbf{X} = 0$ from the previous section is set as the initial value $\mathbf{X}_0$.

- When $\mathbf{X}_0$ is projected onto each camera, it produces $\hat{\mathbf{x}}, \hat{\mathbf{x}}'$, and at this time, minimizing $d(\mathbf{x}, \hat{\mathbf{x}})^2 + d(\mathbf{x}', \hat{\mathbf{x}}')^2$ is aimed, where $d(\mathbf{x}_1, \mathbf{x}_2)$ represents the distance between $\mathbf{x}_1$ and $\mathbf{x}_2$.
- The optimal solution $\mathbf{X}^*$ is obtained iteratively using nonlinear least squares methods such as Gauss-Newton or Levenberg-Marquardt.

Reformulation of the minimization problem



Optimal Triangulation can be performed non-iteratively. This method involves optimizing the distance $d(\mathbf{x}, \mathbf{l}(t))$ between $\mathbf{x}$ and the parameterized Epipolar Line $\mathbf{l}(t)$, instead of optimizing $d(\mathbf{x}, \hat{\mathbf{x}})$. Thus, the optimal solution is obtained by finding the parameter t that minimizes the distance between $\mathbf{x}$ and $\mathbf{l}(t)$.

$$\min_t d(\mathbf{x}, \mathbf{l}(t))^2 + d(\mathbf{x}', \mathbf{l}'(t))^2 \tag{326}$$

Details of the minimization

Initially, $\mathbf{x} = (x, y, 1)^\intercal$ and $\mathbf{x}' = (x', y', 1)^\intercal$ are both transformed to the origin $(0, 0, 1)^\intercal$.

$$\begin{aligned} \mathbf{x} &= (x, y, 1)^\intercal \rightarrow (0, 0, 1)^\intercal \\ \mathbf{x}' &= (x', y', 1)^\intercal \rightarrow (0, 0, 1)^\intercal \end{aligned} \tag{327}$$

The transformation matrix is set as

$$\mathbf{T} = \begin{bmatrix} 1 & & -x \\ & 1 & -y \\ & & 1 \end{bmatrix} \quad \mathbf{T}' = \begin{bmatrix} 1 & & -x' \\ & 1 & -y' \\ & & 1 \end{bmatrix} \tag{328}$$

Next, Epipole $\mathbf{e} = (e_1, e_2, e_3)^\intercal$ and $\mathbf{e}' = (e_1', e_2', e_3')^\intercal$ are each transformed into points on the x-axis, $(1, 0, f)^\intercal$ and $(1, 0, f')^\intercal$ respectively.

$$\begin{aligned} \mathbf{e} &= (e_1, e_2, e_3)^\intercal \rightarrow (1, 0, f)^\intercal \\ \mathbf{e}' &= (e_1', e_2', e_3')^\intercal \rightarrow (1, 0, f')^\intercal \end{aligned} \tag{329}$$

These correspond to $(1/f, 0)^\intercal$ and $(1/f', 0)^\intercal$ in the image plane. After normalizing the Epipole so that $e_1^2 + e_2^2 = e_1'^2 + e_2'^2 = 1$, matrices $\mathbf{R}$ and $\mathbf{R}'$ are defined to rotate this to points on the x-axis.

$$\mathbf{R} = \begin{bmatrix} e_1 & e_2 & \\ -e_2 & e_1 & \\ & & 1 \end{bmatrix} \quad \mathbf{R}' = \begin{bmatrix} e_1' & e_2' & \\ -e_2' & e_1' & \\ & & 1 \end{bmatrix} \tag{330}$$

This transforms $\mathbf{Re} = (1, 0, e_3)^\intercal$ and $\mathbf{R}'\mathbf{e}' = (1, 0, e_3')^\intercal$, where $e_3 = f$ and $e_3' = f'$. Next, the Epipolar Line $\mathbf{l}$ is parameterized to $\mathbf{l}(t)$. As the Epipole $\mathbf{e} = (1, 0, f)^\intercal$ lies on the x-axis, the Epipolar Line passing through it can be parameterized as $(0, t, 1)^\intercal$ relative to the y-axis. Thus $\mathbf{l}(t)$ becomes

$$
\begin{aligned}
\mathbf{l}(t) &= \begin{pmatrix} 0 \\ t \\ 1 \end{pmatrix} \times \begin{pmatrix} 1 \\ 0 \\ f \end{pmatrix} \\
&= \begin{pmatrix} tf \\ 1 \\ -t \end{pmatrix}
\end{aligned} \tag{331}
$$

Since $\mathbf{x}$ has been moved to the origin $(0, 0, 1)^\intercal$ in the previous step, $d(\mathbf{x}, \mathbf{l}(t))$ becomes

$$
d(\mathbf{x}, \mathbf{l}(t)) = \frac{t^2}{1^2 + (tf)^2} \tag{332}
$$

Next, $\mathbf{l}'(t)$ is calculated using the Fundamental Matrix $\mathbf{F}$.

$$
\begin{aligned}
\mathbf{l}'(t) = \mathbf{F} \begin{pmatrix} 0 \\ t \\ 1 \end{pmatrix} \\
\because \mathbf{l}(t) \text{ is the epipolar line of } \begin{pmatrix} 0 & t & 1 \end{pmatrix}^\intercal .
\end{aligned} \tag{333}
$$

$\mathbf{F}$ can be calculated from the given camera matrix pair using $\mathbf{F}_0$, and then calculated from $\mathbf{F}_0$ using $\mathbf{T}, \mathbf{R}, \mathbf{T}', \mathbf{R}'$. The relationship between the Epipole and $\mathbf{F}$ holds that

$$
\begin{aligned}
\mathbf{F} \begin{pmatrix} 1 \\ 0 \\ f \end{pmatrix} = 0 \\
\begin{pmatrix} 1 & 0 & f' \end{pmatrix} \mathbf{F} = 0
\end{aligned} \tag{334}
$$

Thus, $\mathbf{F}_{1,col} = -f\mathbf{F}_{3,col}$ and $\mathbf{F}_{1,row} = -f'\mathbf{F}_{3,row}$, which leads to

$$
\mathbf{F} = \begin{bmatrix} f'fd & -f'c & -f'd \\ -fb & a & b \\ -fd & c & d \end{bmatrix} \tag{335}
$$

Through this, $\mathbf{l}'(t)$ is re-expressed as

$$
\begin{aligned}
\mathbf{l}'(t) = \mathbf{F} \begin{pmatrix} 0 \\ t \\ 1 \end{pmatrix} &= t\mathbf{F}_{2,col} + \mathbf{F}_{3,col} \\
&= \begin{pmatrix} -f'(ct + d) \\ at + b \\ ct + d \end{pmatrix}^\intercal
\end{aligned} \tag{336}
$$

Next, the distance $d(\mathbf{x}', \mathbf{l}'(t))$ between the origin $\mathbf{x}'$ and $\mathbf{l}'(t)$ is calculated as

$$
d(\mathbf{x}', \mathbf{l}'(t))^2 = \frac{(ct + d)^2}{(at + b)^2 + f'^2(ct + d)^2} \tag{337}
$$

Through this, the objective function $d(\mathbf{x}, \mathbf{l}(t))^2 + d(\mathbf{x}', \mathbf{l}'(t))^2$ to be optimized becomes

$$
s(t) = \frac{t^2}{1 + f^2t^2} + \frac{(ct + d)^2}{(at + b)^2 + f'^2(ct + d)^2} \tag{338}
$$

Differentiating this and finding t where $s'(t) = 0$, yields six values of t_i, $i = 1, \ldots, 6$, and the minimum $s(t_i)$ value among them is selected. Using this $t_{\min}$, the optimal Epipolar Lines $\mathbf{l}(t_{\min}), \mathbf{l}'(t_{\min})$ are determined. Next,

$$
\begin{aligned}
\hat{\mathbf{x}} &= \mathbf{l}(t_{\min}) \\
\hat{\mathbf{x}}' &= \mathbf{l}'(t_{\min}) \\
\hat{\mathbf{x}} &\leftarrow \mathbf{T}^{-1}\mathbf{R}^{-1}\hat{\mathbf{x}} \\
\hat{\mathbf{x}}' &\leftarrow \mathbf{T}'^{-1}\mathbf{R}'^{-1}\hat{\mathbf{x}}'
\end{aligned} \tag{339}
$$

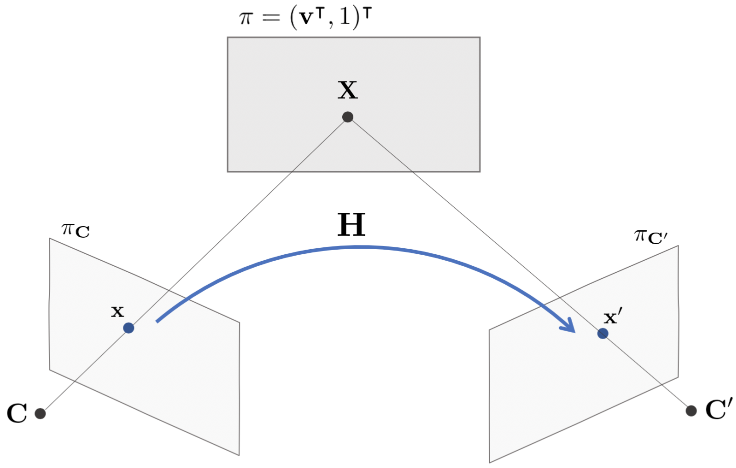
Following this, the original $\hat{\mathbf{x}}, \hat{\mathbf{x}}'$ are restored and then

$$\begin{aligned}\hat{\mathbf{x}} \times \mathbf{P}\hat{\mathbf{X}} &= 0 \\ \hat{\mathbf{x}}' \times \mathbf{P}\hat{\mathbf{X}}' &= 0\end{aligned} \tag{340}$$

The equations are reformulated to the form $\mathbf{A}\hat{\mathbf{X}} = 0$, and the optimal solution $\hat{\mathbf{X}}$ is finally obtained through Singular Value Decomposition (SVD). This method of finding an approximate solution for $\mathbf{X}$ is called the Optimal Triangulation method.

10 Scene planes and homographies

Homographies given the plane and vice versa



This section explains the homography transformation from one image plane to another using a plane π in the world.

Result 13.1

When the pairs of camera matrices $(\mathbf{P}, \mathbf{P}')$ transformed into Canonical Form and the plane π in the world are given as follows:

$$\begin{aligned}&\mathbf{P} = [\mathbf{I} \mid 0] \\ &\mathbf{P}' = [\mathbf{A} \mid \mathbf{a}] \\ &\pi = (\mathbf{v}^\intercal, 1)^\intercal \quad \mathbf{v} \in \mathbb{R}^3\end{aligned} \tag{341}$$

Then, the Homography $\mathbf{H}$ is given by

$$\mathbf{H} = \mathbf{A} - \mathbf{a}\mathbf{v}^\intercal \tag{342}$$

Proof

When the image planes $\pi_{\mathbf{P}}, \pi_{\mathbf{P}'}$ of the two cameras are given, the point $\mathbf{X} \in \pi$ on the world plane is projected to the point $\mathbf{x}$ on the plane $\pi_{\mathbf{P}}$ as follows.

$$\begin{aligned}\mathbf{P}\mathbf{X} &= [\mathbf{I} \mid 0] \begin{bmatrix} \tilde{\mathbf{X}} \\ 1 \end{bmatrix} \\ &= \tilde{\mathbf{X}} = \mathbf{x}\end{aligned} \tag{343}$$

Therefore, for an arbitrary scalar ρ, $\tilde{\mathbf{X}} = \rho^{-1}\mathbf{x}$ holds. Rewriting $\mathbf{X}$, $\mathbf{X} = \begin{bmatrix} x & \rho \end{bmatrix}^\intercal$ and since $\mathbf{X}$ is a point on the plane π,

$$\begin{bmatrix} \mathbf{v}^\intercal & 1 \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \rho \end{bmatrix} = 0 \tag{344}$$

holds. Solving this gives $\rho = -\mathbf{v}^\intercal \mathbf{x}$. Projecting $\mathbf{X}$ onto $\pi_{\mathbf{P}'}$ results in

$$\begin{aligned} \mathbf{x}' = \mathbf{P}'\mathbf{X} &= [\mathbf{A} \mid \mathbf{a}] \begin{bmatrix} \mathbf{x} \\ -\mathbf{v}^\intercal \mathbf{x} \end{bmatrix} \\ &= \mathbf{A}\mathbf{x} - \mathbf{a}\mathbf{v}^\intercal \mathbf{x} \\ &= (\mathbf{A} - \mathbf{a}\mathbf{v}^\intercal)\mathbf{x} \end{aligned} \tag{345}$$

Therefore, $\mathbf{H} = \mathbf{A} - \mathbf{a}\mathbf{v}^\intercal$.

A calibrated stereo rig

If the pair of calibrated stereo camera matrices $(\mathbf{P}, \mathbf{P}')$ is given as follows and

$$\begin{aligned} \mathbf{P} &= \mathbf{K}[\mathbf{I} \mid 0] \\ \mathbf{P}' &= \mathbf{K}'[\mathbf{R} \mid \mathbf{t}] \end{aligned} \tag{346}$$

a plane $\pi = (\mathbf{n}^\intercal, d)^\intercal$ in the world is given, then the condition

$$\mathbf{n}^\intercal \tilde{\mathbf{X}} + d = 0 \tag{347}$$

holds. This can be rearranged to $-\frac{\mathbf{n}^\intercal \tilde{\mathbf{X}}}{d} = 1$. The point $\mathbf{x}$ projected onto the image plane $\pi_{\mathbf{P}}$ of $\mathbf{X}$ can be expressed as follows.

$$\begin{aligned} \mathbf{x} = \mathbf{P}\mathbf{X} &= \mathbf{K}[\mathbf{I} \mid 0] \begin{bmatrix} \tilde{\mathbf{X}} \\ 1 \end{bmatrix} \\ &= \mathbf{K}\tilde{\mathbf{X}} \end{aligned} \tag{348}$$

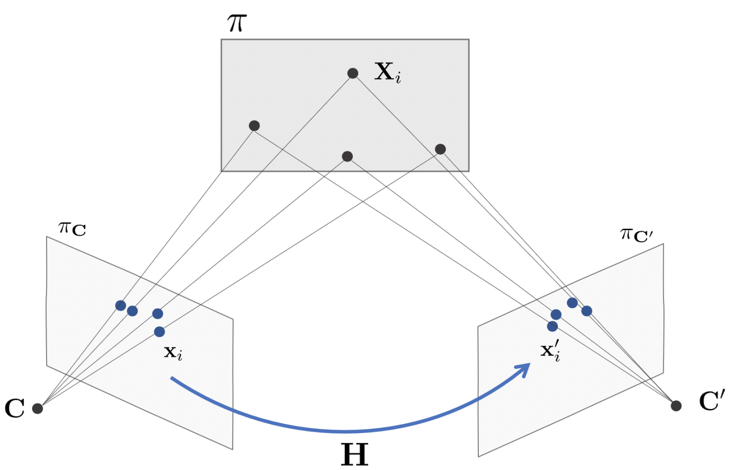
Therefore, $\tilde{\mathbf{X}} = \mathbf{K}^{-1}\mathbf{x}$, and next, the point $\mathbf{x}'$ projected onto the image plane $\pi_{\mathbf{P}'}$ of $\mathbf{X}$ is

$$\begin{aligned} \mathbf{x}' = \mathbf{P}'\mathbf{X} &= \mathbf{K}'[\mathbf{R} \mid \mathbf{t}] \begin{bmatrix} \tilde{\mathbf{X}} \\ 1 \end{bmatrix} \\ &= \mathbf{K}'\mathbf{R}\tilde{\mathbf{X}} + \mathbf{K}'\mathbf{t}(-\frac{\mathbf{n}^\intercal \tilde{\mathbf{X}}}{d}) \\ &= \mathbf{K}'(\mathbf{R} - \frac{\mathbf{t}\mathbf{n}^\intercal}{d})\tilde{\mathbf{X}} \\ &= \mathbf{K}'(\mathbf{R} - \frac{\mathbf{t}\mathbf{n}^\intercal}{d})\mathbf{K}^{-1}\mathbf{x} \end{aligned} \tag{349}$$

Therefore, the Homography $\mathbf{H}$ for a calibrated stereo camera becomes

$$\mathbf{H} = \mathbf{K}'(\mathbf{R} - \frac{\mathbf{t}\mathbf{n}^\intercal}{d})\mathbf{K}^{-1} \tag{350}$$

Homographies compatible with epipolar geometry



Assuming that there are four corresponding point pairs $(\mathbf{x}_i, \mathbf{x}'_i), \ i = 1, \cdots, 4$ existing on the plane π in the world. We can obtain eight constraints from four pairs of corresponding points, which uniquely determine the Homography $\mathbf{H}$.

Next, let's assume that four arbitrary corresponding point pairs $(\mathbf{x}_i, \mathbf{x}'_i), i = 1, \cdots, 4$ not collinear (no three are collinear) are given in the world. **In this case, a Homography H exists that transforms the points from one image plane $\pi_\mathbf{P}$ to another image plane $\pi_{\mathbf{P}'}$. For this H to be compatible with the Fundamental Matrix F between the two cameras, it must be a Homography transformation concerning the plane π in the world. In other words, for H to follow the Epipolar Geometry, it must be a Homography concerning π.** The condition for $\mathbf{x}$ to correspond with $\mathbf{Hx}$ according to Epipolar Geometry is

$$(\mathbf{Hx})^\intercal \mathbf{Fx} = 0 \tag{351}$$

Thus, $\mathbf{x}^\intercal \mathbf{H}^\intercal \mathbf{Fx} = 0$ must be satisfied, meaning $\mathbf{H}^\intercal \mathbf{F}$ must be a skew-symmetric matrix.

$$\mathbf{H}^\intercal \mathbf{F} + \mathbf{F}^\intercal \mathbf{H} = 0 \tag{352}$$

This condition is the necessary and sufficient condition for H to follow the Epipolar Geometry.

Result 13.3

Generally, when $\mathbf{F} = \mathbf{e}'^\wedge \mathbf{A}$ is given, $\mathbf{H}$ is

$$\mathbf{H} = \mathbf{A} - \mathbf{e}'\mathbf{v}^\intercal \tag{353}$$

In this case, $\mathbf{H}$ has three degrees of freedom from $\mathbf{v} \in \mathbb{R}^3$.

Proof

When $\mathbf{F} = \mathbf{e}'^\wedge \mathbf{A}$ is given as such, performing a Projective Reconstruction can determine the pairs of camera matrices as

$$\begin{aligned} \mathbf{P} &= [\mathbf{I} \mid 0] \\ \mathbf{P}' &= [\mathbf{A} \mid \mathbf{e}'] \end{aligned} \tag{354}$$

Given the plane $\pi = (\mathbf{v}^\intercal, 1)^\intercal$ in the world, from the previous theorem, it can be derived as

$$\mathbf{H} = \mathbf{A} - \mathbf{e}'\mathbf{v} \tag{355}$$

Inserting this into $\mathbf{F}^\intercal \mathbf{H}$ results in

$$\begin{aligned} \mathbf{F}^\intercal \mathbf{H} &= -\mathbf{A}^\intercal \mathbf{e}'^\wedge (\mathbf{A} - \mathbf{e}'^\intercal \mathbf{v}) \\ &= -\mathbf{A}^\intercal \mathbf{e}'^\wedge \mathbf{A} \quad \because \mathbf{e}'^\wedge \mathbf{e}' = 0 \end{aligned} \tag{356}$$

which becomes a skew-symmetric matrix.

Corollary

Any Homography $\mathbf{H}$ between two cameras must be a Homography concerning the plane π in the world to follow the Epipolar Geometry. The necessary and sufficient condition for $\mathbf{H}$ to be a Homography concerning the plane π is as follows.

$$\mathbf{F} = \mathbf{e}'^\wedge \mathbf{H} \tag{357}$$

Proof

($\Rightarrow$) If $\mathbf{H}$ is a Homography concerning the plane $\pi = (\mathbf{v}^\intercal, 1)^\intercal$ while $\mathbf{F} = \mathbf{e}'^{\wedge}\mathbf{A}$, then according to the previous theorem,

$$\mathbf{H} = \mathbf{H}_{\begin{pmatrix}\mathbf{v} \\ 1\end{pmatrix}} = \mathbf{A} - \mathbf{e}'^\intercal\mathbf{v} \tag{358}$$

holds. Therefore,

$$\begin{aligned}\mathbf{e}'^{\wedge}\mathbf{H} &= \mathbf{e}'^{\wedge}(\mathbf{A} - \mathbf{e}'^\intercal\mathbf{v}) \\ &= \mathbf{e}'^{\wedge}\mathbf{A} = \mathbf{F} \quad \because \mathbf{e}'^{\wedge}\mathbf{e}' = 0\end{aligned} \tag{359}$$

results.

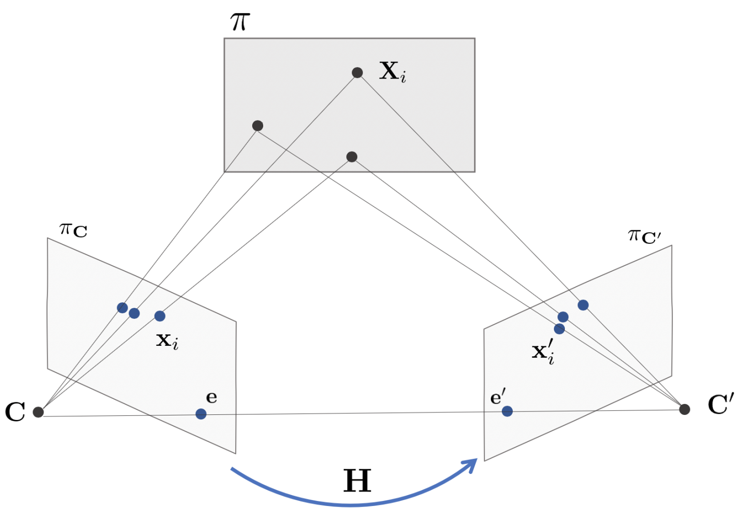
($\Leftarrow$) If $\mathbf{F} = \mathbf{e}'^{\wedge}\mathbf{H}$, then $\mathbf{H}$ is a Homography due to any plane π in the world. For example, if $\mathbf{H}$ is due to the infinite plane $\pi_\infty = (\mathbf{0}^\intercal, 1)^\intercal$,

$$\mathbf{H} = \mathbf{H}_{\begin{pmatrix}\mathbf{0} \\ 1\end{pmatrix}} = \mathbf{H} - \mathbf{e}' \cdot 0 = \mathbf{H} \tag{360}$$

results.

Plane induced homographies given F and image correspondences

Three points



Given three corresponding point pairs $(\mathbf{x}_i, \mathbf{x}'_i),\ i = 1, 2, 3$ on two image planes $\pi_{\mathbf{P}}, \pi_{\mathbf{P}'}$, the points $\mathbf{X}_i,\ i = 1, 2, 3$ in the world obtained by back-projection through these points uniquely determine a plane π in the world, and thus enable the computation of a homography $\mathbf{H}$. This section describes how to compute such a Homography $\mathbf{H}$.

There are two main methods to compute $\mathbf{H}$. **The first method involves directly determining the world plane π.** Given three corresponding point pairs $(\mathbf{x}_i, \mathbf{x}'_i),\ i = 1, 2, 3$ and the Fundamental Matrix $\mathbf{F} = \mathbf{e}'^{\wedge}\mathbf{A}$, one can compute the camera matrix (projective reconstruction). Using the given corresponding pairs $(\mathbf{x}_i, \mathbf{x}'_i),\ i = 1, 2, 3$ and $\mathbf{F}$, we can determine (up to projectivity):

$$\begin{aligned}\mathbf{P} &= [\mathbf{I} \mid 0] \\ \mathbf{P}' &= [\mathbf{A} \mid \mathbf{e}']\end{aligned} \tag{361}$$

Additionally, the world points $\mathbf{X}_i,\ i = 1, 2, 3$ can be computed using $\mathbf{P}, \mathbf{P}'$. These points $\mathbf{X}_i,\ i = 1, 2, 3$ must not be collinear. Next, the plane π that includes the three world points $\mathbf{X}_i,\ i = 1, 2, 3$ uniquely determines $\pi = (\mathbf{v}^\intercal, d)^\intercal$, and thus,

$$\mathbf{H} = \mathbf{A} - \mathbf{e}'\mathbf{v}^\intercal \tag{362}$$

can be used to calculate the Homography $\mathbf{H}$ that transforms points between $\pi_\mathbf{P} \leftrightarrow \pi_{\mathbf{P}'}$ based on π.

The second method involves solving $\mathbf{H}\mathbf{x}_i = \mathbf{x}'_i,\ \ i = 1, 2, 3$ algebraically. Here, the equation $\mathbf{x}'_i \times (\mathbf{H}\mathbf{x}_i) = 0$ is transformed into a linear equation form $\mathbf{A}\mathbf{h} = 0$ to compute Homography $\mathbf{H}$. However, to compute Homography $\mathbf{H}$, a total of four corresponding point pairs are needed, but only three are provided, thus an additional corresponding point pair is required. Using the known Fundamental Matrix $\mathbf{F}$, one can find the epipoles $\mathbf{e}, \mathbf{e}'$ of the two image planes and back-project them to add one more corresponding point pair to compute Homography $\mathbf{H}$. It is assumed here that the point pairs $(\mathbf{x}_i, \mathbf{x}'_i), i = 1, 2, 3$ must not reside on the Epipolar Line.

Result 13.6

Given the Fundamental Matrix $\mathbf{F}$ and three corresponding point pairs $(\mathbf{x}_i, \mathbf{x}'_i), i = 1, 2, 3$, the Homography $\mathbf{H}$ induced by the computed world plane π can be calculated as:

$$\mathbf{H} = \mathbf{A} - \mathbf{e}'(\mathbf{M}^{-1}\mathbf{b})^\intercal \tag{363}$$

Here,

$$\begin{aligned} \mathbf{M} &= \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \end{bmatrix} \\ \mathbf{b} &= (\mathbf{x}'_i \times (\mathbf{A}\mathbf{x}_i))^\intercal (\mathbf{x}'_i \times \mathbf{e}') / \|\mathbf{x}'_i \times \mathbf{e}'\| \end{aligned} \tag{364}$$

Proof

When the Canonical camera matrix $\mathbf{P} = [\mathbf{I} \mid 0], \mathbf{P}' = [\mathbf{A} \mid \mathbf{e}']$ is given, the Fundamental Matrix $\mathbf{F} = \mathbf{e}'^\wedge \mathbf{A}$ can be computed, resulting in

$$\mathbf{e}'^\wedge \mathbf{F} = \mathbf{e}'^\wedge \mathbf{e}'^\wedge \mathbf{A} \sim \mathbf{A} \tag{365}$$

showing a proportional relationship with $\mathbf{A}$, thus $\mathbf{A} = \mathbf{e}'^\wedge \mathbf{F}$ holds.

When a world plane $\pi = (\mathbf{v}^\intercal, d)^\intercal$ is given, any Homography $\mathbf{H}$ can be expressed as $\mathbf{H} = \mathbf{A} - \mathbf{e}'\mathbf{v}$, and to find $\mathbf{v}$,

$$\mathbf{x}'_i \times \mathbf{H}\mathbf{x}_i = 0 \tag{366}$$

is set up and rearranged to:

$$\begin{aligned} \mathbf{x}'_i \times \mathbf{H}\mathbf{x}_i &= 0 \\ \mathbf{x}'_i \times (\mathbf{A}\mathbf{x}_i - \mathbf{e}'\mathbf{v}^\intercal \mathbf{x}_i) &= 0 \\ \mathbf{x}'_i \times \mathbf{A}\mathbf{x}_i &= (\mathbf{x}_i \times \mathbf{e}')\mathbf{v}^\intercal \mathbf{x}_i \end{aligned} \tag{367}$$

Multiplying both sides by $(\mathbf{x}_i \times \mathbf{e}')^\intercal$ and rearranging gives:

$$\mathbf{x}_i^\intercal \mathbf{v} = \mathbf{b}_i \tag{368}$$

Applying this to all three points results in:

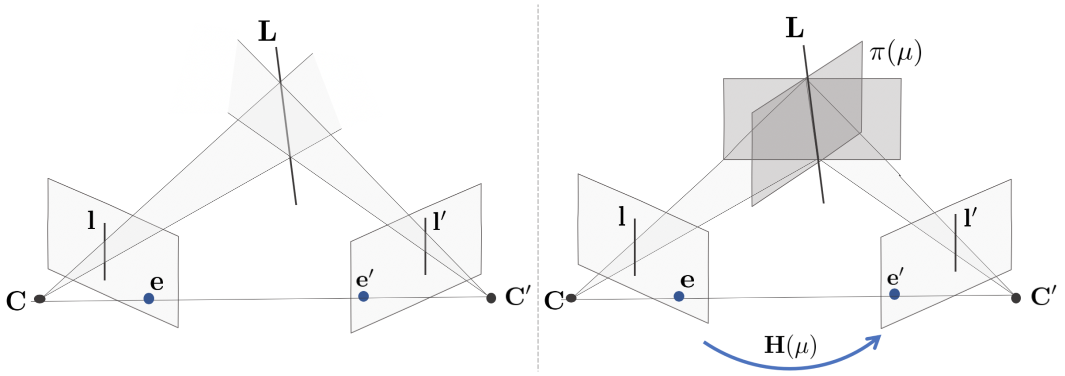
$$\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \end{bmatrix} \mathbf{v} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \mathbf{b}_3 \end{bmatrix} \tag{369}$$

Substituting $\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \end{bmatrix}$ with matrix $\mathbf{M}$ leads to:

$$\mathbf{M}\mathbf{v} = \mathbf{b} \tag{370}$$

Rearranging this results in **$\mathbf{v} = \mathbf{M}^{-1}\mathbf{b}$, leading to the Homography $\mathbf{H} = \mathbf{A} - \mathbf{e}'(\mathbf{M}^{-1}\mathbf{b})^\intercal$ as shown.**

A point and line



This section explains how to compute a Homography $\mathbf{H}$ given a line and a point in the world.

Result 13.7

When corresponding image plane lines $\mathbf{l} \leftrightarrow \mathbf{l}'$ for a world line $\mathbf{L}$ are given, the Homography $\mathbf{H}$ that transforms $\mathbf{l}$ to $\mathbf{l}'$ using the world line $\mathbf{L}$ can be represented as:

$$\mathbf{H}(\mu) = \mathbf{l}'^{\wedge}\mathbf{F} + \mu \mathbf{e}'\mathbf{l}^{\intercal} \tag{371}$$

Here, $\mathbf{l}'^{\intercal}\mathbf{e}' \neq 0$ must hold, and μ is a parameter where $\mu \in \mathbb{P}^1$.

Proof

Given the corresponding camera matrices $\mathbf{P} = [\mathbf{I} \mid 0], \mathbf{P}' = [\mathbf{A} \mid \mathbf{e}']$, the planes $\pi_{\mathbf{l}} = \mathbf{P}^{\intercal}\mathbf{l}, \pi_{\mathbf{l}'} = \mathbf{P}'^{\intercal}\mathbf{l}'$ projected back from image plane lines $\mathbf{l}, \mathbf{l}'$ can be represented. The intersecting line of these two back-projected planes $\pi_{\mathbf{l}}$, $\pi_{\mathbf{l}'}$ forms the world line $\mathbf{L}$, and the plane π containing $\mathbf{L}$ can be parameterized by μ as:

$$\begin{aligned} \pi(\mu) &= \mu \mathbf{P}^{\intercal}\mathbf{l} + \mathbf{P}'^{\intercal}\mathbf{l}' \\ &= \mu \begin{pmatrix} \mathbf{l} \\ 0 \end{pmatrix} + \begin{pmatrix} \mathbf{A}^{\intercal}\mathbf{l}' \\ \mathbf{e}'^{\intercal}\mathbf{l}' \end{pmatrix} \\ &= \begin{pmatrix} \mu\mathbf{l} + \mathbf{A}^{\intercal}\mathbf{l}' \\ \mathbf{e}'^{\intercal}\mathbf{l}' \end{pmatrix} \end{aligned} \tag{372}$$

From Result 13.1, Homography $\mathbf{H}$ using the plane $\pi(\mu)$ can be expressed as:

$$\mathbf{H}(\mu) = \mathbf{A} - \mathbf{e}'\mathbf{v}(\mu)^{\intercal} \tag{373}$$

Here, $\mathbf{v}$ according to the formula for $\pi(\mu)$ above becomes $\mathbf{v}(\mu) = (\mu\mathbf{l} + \mathbf{A}^{\intercal}\mathbf{l}')/(\mathbf{e}'^{\intercal}\mathbf{l}')$. Substituting the matrix $\mathbf{A}$ with $\mathbf{A} = \mathbf{e}'^{\wedge}\mathbf{F}$ and rearranging gives:

$$\mathbf{H}(\mu) = -(\mathbf{l}'^{\wedge}\mathbf{F} + \mu\mathbf{e}'\mathbf{l}^{\intercal})/(\mathbf{e}'^{\intercal}\mathbf{l}') \sim \mathbf{l}'^{\wedge}\mathbf{F} + \mu\mathbf{e}'\mathbf{l}^{\intercal} \tag{374}$$

Thus, **H is thus proportional to $\mathbf{H}(\mu) = \mathbf{l}'^{\wedge}\mathbf{F} + \mu\mathbf{e}'\mathbf{l}^{\intercal}$, so it can be written as this equation.**

The homography for a corresponding point and line

The Homography $\mathbf{H}(\mu)$ using the corresponding line pairs $\mathbf{l} \leftrightarrow \mathbf{l}'$ varies depending on the μ value.

Result 13.8

Given one corresponding line pair $\mathbf{l} \leftrightarrow \mathbf{l}'$ and one corresponding point pair $\mathbf{x} \leftrightarrow \mathbf{x}'$, a unique Homography $\mathbf{H}$ can be derived as follows:

$$\mathbf{H} = \mathbf{l}'^{\wedge}\mathbf{F} + \frac{(\mathbf{x}' \times \mathbf{e}')^{\intercal}(\mathbf{x}' \times ((\mathbf{F}\mathbf{x}) \times \mathbf{l}'))}{\|\mathbf{x}' \times \mathbf{e}'\|^2(\mathbf{l}^{\intercal}\mathbf{x})}\mathbf{e}'\mathbf{l}^{\intercal} \tag{375}$$

Using the corresponding line pair $\mathbf{l} \leftrightarrow \mathbf{l}'$, Homography $\mathbf{H}(\mu) = \mathbf{l}'^{\wedge}\mathbf{F} + \mu\mathbf{e}'\mathbf{l}^{\intercal}$ can be calculated. When expanding the equation $\mathbf{Hx} = \mathbf{x}'$ to $\mathbf{x}' \times (\mathbf{Hx}) = 0$, we get:

$$\mathbf{x}' \times (\mathbf{l}'^{\wedge}\mathbf{Fx} + \mu\mathbf{e}'\mathbf{l}^{\intercal}\mathbf{x}) = 0 \tag{376}$$

Upon expansion and rearrangement:

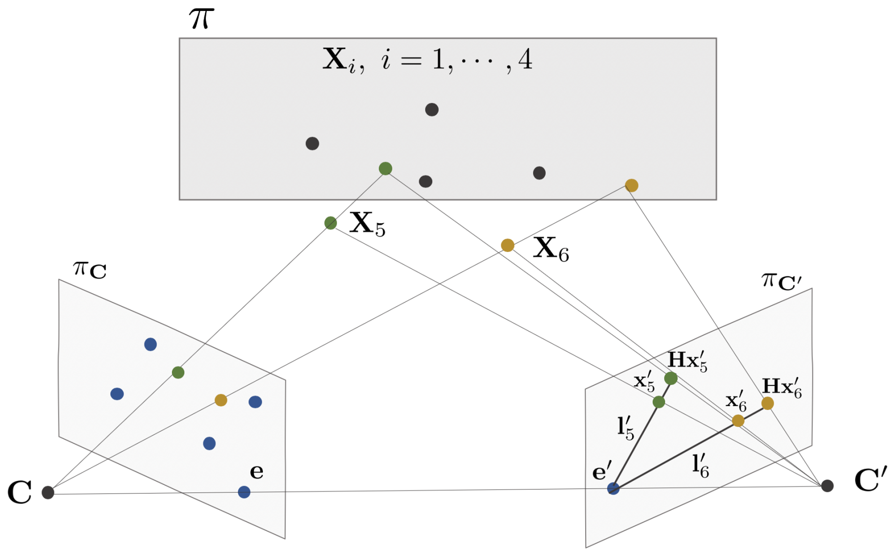
$$\begin{aligned}(\mathbf{x}' \times \mathbf{e}')\mathbf{l}^{\intercal}\mathbf{x} \cdot \mu &= -\mathbf{x}' \times \mathbf{l}' \times \mathbf{Fx} \\ &= \mathbf{x}' \times \mathbf{Fx} \times \mathbf{l}'\end{aligned} \tag{377}$$

After multiplying by $(\mathbf{x}' \times \mathbf{e}')^{\intercal}$ and rearranging for μ, then substituting into $\mathbf{H}(\mu)$ equation, we obtain:

$$\mathbf{H} = \mathbf{l}'^{\wedge}\mathbf{F} + \frac{(\mathbf{x}' \times \mathbf{e}')^{\intercal}(\mathbf{x}' \times ((\mathbf{Fx}) \times \mathbf{l}'))}{\|\mathbf{x}' \times \mathbf{e}'\|^2(\mathbf{l}^{\intercal}\mathbf{x})}\mathbf{e}'\mathbf{l}^{\intercal} \tag{378}$$

The formula is derived as shown.

Computing F given the homography induced by a plane



Generally, to compute the Fundamental Matrix $\mathbf{F}$ for a pair of camera matrices $(\mathbf{P}, \mathbf{P}')$, at least 8 corresponding point pairs $(\mathbf{x}_i, \mathbf{x}'_i),\ i = 1, \cdots, 8$ are required. However, **by using Scene Plane Homography, F can be computed using only 6 corresponding point pairs.** There is a constraint that the world points of 4 of these corresponding point pairs $(\mathbf{x}_i, \mathbf{x}'_i),\ i = 1, \cdots, 4$ must exist on the same plane π.

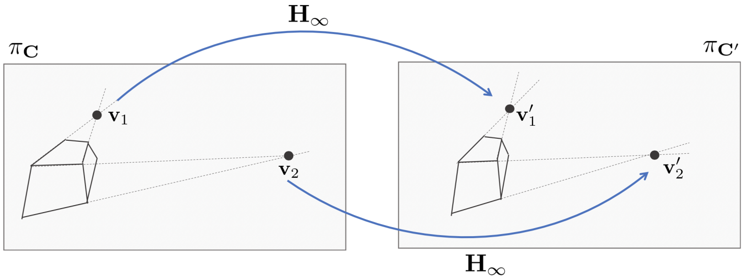
The algorithm for computing the Fundamental Matrix $\mathbf{F}$ with 6 corresponding point pairs is as follows:

- Use the 4 point pairs $(\mathbf{x}_i, \mathbf{x}'_i),\ i = 1, \cdots, 4$, projected from the world plane π, to compute the Homography $\mathbf{H}$ such that $\mathbf{x}'_i = \mathbf{H}\mathbf{x}_i$. These 4 pairs of corresponding points uniquely determine the Homography $\mathbf{H}$.
- Use the remaining 2 corresponding point pairs $(\mathbf{x}_i, \mathbf{x}'_i),\ i = 5, 6$ to determine the Epipole $\mathbf{e}'$. Use $\mathbf{H}\mathbf{x}_5$, the homography transformation of $\mathbf{x}_5$, and $\mathbf{x}'_5$ to form a line as $(\mathbf{H}\mathbf{x}_5) \times \mathbf{x}'_5$. Do the same for $\mathbf{x}_6$ and find the intersection point of the two lines, which will be the Epipole $\mathbf{e}'$.

$$\mathbf{e}' = (\mathbf{H}\mathbf{x}_5) \times \mathbf{x}'_5 \cap (\mathbf{H}\mathbf{x}_6) \times \mathbf{x}'_6 \tag{379}$$

- Compute the Fundamental Matrix $\mathbf{F}$ through $\mathbf{F} = \mathbf{e}'^{\wedge}\mathbf{H}$.

The infinite homography $\mathbf{H}_\infty$



Definition 13.10

When the Scene Plane is the plane at infinity π_∞, the calculated Homography is called the Infinity Homography $\mathbf{H}_\infty$.

When the pair of camera matrices $(\mathbf{P}, \mathbf{P}')$ are given as follows

$$\mathbf{P} = \mathbf{K}[\mathbf{I} \mid 0] \quad \mathbf{P}' = \mathbf{K}'[\mathbf{R} \mid \mathbf{t}] \tag{380}$$

The Homography for the world plane $\pi(d) = \begin{pmatrix} \mathbf{v} \\ d \end{pmatrix}$ can be calculated as follows:

$$\mathbf{H}_{\pi(d)} = \mathbf{K}'(\mathbf{R} - \frac{\mathbf{t}\mathbf{n}^\intercal}{d})\mathbf{K}^{-1} \tag{381}$$

Infinity Homography $\mathbf{H}_\infty$ corresponds to the case where d is infinite, therefore

$$\mathbf{H}_\infty = \lim_{d\to\infty} \mathbf{H}_{\pi(d)} = \mathbf{K}'\mathbf{R}\mathbf{K}^{-1} \tag{382}$$

Thus, **it can be seen that $\mathbf{H}_\infty$ depends only on the camera's rotation, not on $\mathbf{t}$.**

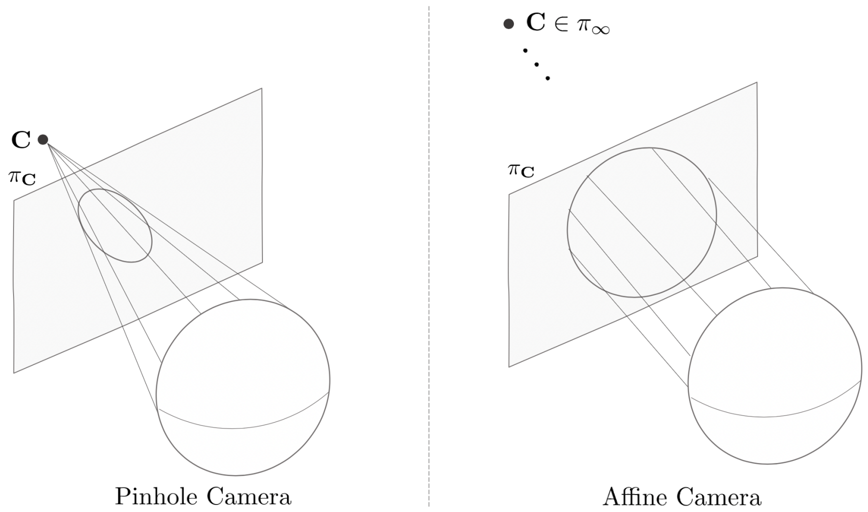
Affine and metric reconstruction

If the world coordinate system is Affine, π_∞ can be represented as $\pi_\infty = (0,0,0,1)^\intercal$, and if the two camera matrices are given as $\mathbf{P} = [\mathbf{M} \mid \mathbf{m}]$, $\mathbf{P}' = [\mathbf{M}' \mid \mathbf{m}']$, then projecting a point $\mathbf{X} \in \pi_\infty$ onto each camera results in

$$\begin{aligned} \mathbf{x} &= \mathbf{P}\mathbf{X} = [\mathbf{M} \mid \mathbf{m}] \begin{bmatrix} \tilde{\mathbf{X}} \\ 0 \end{bmatrix} = \mathbf{M}\tilde{\mathbf{X}} \\ \mathbf{x}' &= \mathbf{P}'\mathbf{X} = [\mathbf{M}' \mid \mathbf{m}'] \begin{bmatrix} \tilde{\mathbf{X}} \\ 0 \end{bmatrix} = \mathbf{M}'\tilde{\mathbf{X}} = (\mathbf{M}'\mathbf{M}^{-1})\mathbf{x} \end{aligned} \tag{383}$$

Therefore, $\mathbf{x}' = (\mathbf{M}'\mathbf{M}^{-1})\mathbf{x}$, so when the world coordinate system is Affine, the Infinite Homography between the two cameras is $\mathbf{H}_\infty = \mathbf{M}^{'}\mathbf{M}^{-1}$.

11 Affine Epipolar Geometry



A pinhole camera (pinhole camera) refers to a camera model where objects in the world are projected onto the image plane through a single point called the focus. Therefore, the pinhole camera model can be modeled as a mapping function from projective space $\mathbb{P}^3 \mapsto \mathbb{P}^2$, resulting in distortion due to perspective. In contrast, **an Affine camera means a camera where the projection of objects in the world onto the image plane is like the shadow formed by an infinite light source. Thus, an Affine camera does not produce distortion due to perspective.**

An Affine camera can also be represented in Canonical Form. If two Affine cameras $\mathbf{P}_A, \mathbf{P}'_A$ are given when the world coordinate system and the camera coordinate system are the same, the principal axis of $\mathbf{P}_A$ is the Z axis, and the projection is to the XY plane. $\mathbf{P}'_A$ can be represented through an Affine transformation $\mathbf{M}$ and camera movement $\mathbf{t}$. **At this time, both Affine cameras have the property that the last row is $(0, 0, 0, 1)$.**

$$\mathbf{P}_A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \mathbf{P}'_A = \begin{bmatrix} \mathbf{M} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} \quad \text{where, } \mathbf{M} \in \mathbb{R}^{2\times 3}, \ \mathbf{t} \in \mathbb{R}^2 \tag{384}$$

The Affine camera matrix $\mathbf{P}_A$, due to the properties of the Affine transformation, transforms a point on the plane at infinity $\pi_\infty = (X, Y, Z, 0)^\intercal$ into a point on the line at infinity $\mathbf{l}_\infty$, preserving the parallel properties of the object in the world.

$$\mathbf{P}_A \begin{bmatrix} X \\ Y \\ Z \\ 0 \end{bmatrix} = \begin{bmatrix} * \\ * \\ 0 \end{bmatrix} \in \mathbf{l}_\infty \tag{385}$$

In the general Affine camera matrix $\begin{bmatrix} \mathbf{M} & \mathbf{m} \\ \mathbf{0} & 1 \end{bmatrix} \in \mathbb{R}^{3\times 4}$, the matrix $\mathbf{M} \in \mathbb{R}^{2\times 3}$ has rank 2, thus

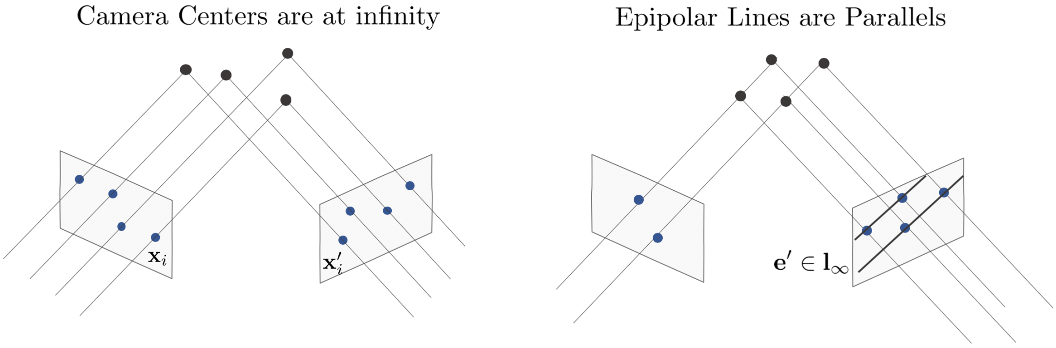
$$\mathbf{M}\tilde{\mathbf{C}} = 0 \tag{386}$$

satisfies the existence of an Affine camera center point $\mathbf{C}$. **At this time, C exists on the plane at infinity.**

$$\mathbf{C} = \begin{pmatrix} \tilde{\mathbf{C}} \\ 0 \end{pmatrix} \in \pi_\infty \tag{387}$$

The center point $\mathbf{C}$ of the Affine camera matrix $\mathbf{P}_A$ is $\begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix}^\intercal$. **That is, the center point of the Affine camera is in the same direction as the principal axis.**

Affine epipolar geometry



Epipolar lines

Given two identical Affine cameras, consider their Back-projection lines. Back-projecting a point $\mathbf{x}$ on the image plane of the first Affine camera results in

$$\mathbf{X}(\lambda) = \mathbf{P}_A^{\dagger}\mathbf{x} + \lambda \begin{bmatrix} \tilde{\mathbf{C}} \\ 0 \end{bmatrix} \tag{388}$$

Therefore, the direction of the Back-projection lines is exactly the direction of the center point of the Affine camera $\begin{bmatrix} \tilde{\mathbf{C}} \\ 0 \end{bmatrix}$. **That is, all the Back-projection lines of the points on the image plane of the Affine camera are parallel, so the Epipolar Lines projected onto the second Affine camera are also parallel.**

The epipoles

Since all Epipolar Lines are parallel, **the Epipoles exist on the line at infinity (line at infinity).**

The affine fundamental matrix

Result 14.1

Given two identical Affine cameras, the Affine Fundamental Matrix $\mathbf{F}_A$ can be defined. At this time, $\mathbf{F}_A$ appears as

$$\mathbf{F}_A = \begin{bmatrix} 0 & 0 & * \\ 0 & 0 & * \\ * & * & * \end{bmatrix} \tag{389}$$

where $*$ means non-zero values. Generally, $\mathbf{F}_A$ is written as

$$\mathbf{F}_A = \begin{bmatrix} 0 & 0 & a \\ 0 & 0 & b \\ c & d & e \end{bmatrix} \tag{390}$$

and has rank 2 like the general Fundamental Matrix.

Derivation

Geometric derivation

Given two Affine camera matrices $\mathbf{P}_A, \mathbf{P}'_A$, parallel lines are preserved due to the properties of Affine transformation. Therefore, the Homography $\mathbf{H}_A$ between the two image planes becomes an Affine Homography, satisfying

$$\mathbf{x}' = \mathbf{H}_A\mathbf{x} \quad \text{where, } \mathbf{H}_A = \begin{bmatrix} * & * & * \\ * & * & * \\ 0 & 0 & 1 \end{bmatrix} \tag{391}$$

Since the Epipole $\mathbf{e}'$ of the Affine camera exists on the line at infinity, its Cross Product appears as

$$\mathbf{e}'^{\wedge} = \begin{bmatrix} 0 & 0 & * \\ 0 & 0 & * \\ * & * & 0 \end{bmatrix} \tag{392}$$

Thus, the Affine Fundamental Matrix $\mathbf{F}_A$ can be calculated through $\mathbf{F}_A = \mathbf{e}'^{\wedge}\mathbf{H}_A$, resulting in

$$\begin{aligned} \mathbf{F}_A &= \mathbf{e}'^{\wedge}\mathbf{H}_A \\ &= \begin{bmatrix} 0 & 0 & * \\ 0 & 0 & * \\ * & * & 0 \end{bmatrix} \begin{bmatrix} * & * & * \\ * & * & * \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & * \\ 0 & 0 & * \\ * & * & * \end{bmatrix} \end{aligned} \tag{393}$$

Properties

The epipoles

Given the Affine Fundamental Matrix $\mathbf{F}_A$, the Epipoles $\mathbf{e}, \mathbf{e}'$ can be calculated. When $\mathbf{F}_A$ is

$$\mathbf{F}_A = \begin{bmatrix} 0 & 0 & a \\ 0 & 0 & b \\ c & d & e \end{bmatrix} \tag{394}$$

due to the properties of the Fundamental Matrix, $\mathbf{F}_A\mathbf{e} = 0$ and $\mathbf{e}'^\intercal\mathbf{F}_A = 0$ hold, allowing the following calculations:

$$\begin{aligned} \mathbf{e} &= \begin{bmatrix} -d & c & 0 \end{bmatrix}^\intercal \\ \mathbf{e}' &= \begin{bmatrix} -b & a & 0 \end{bmatrix}^\intercal \end{aligned} \tag{395}$$

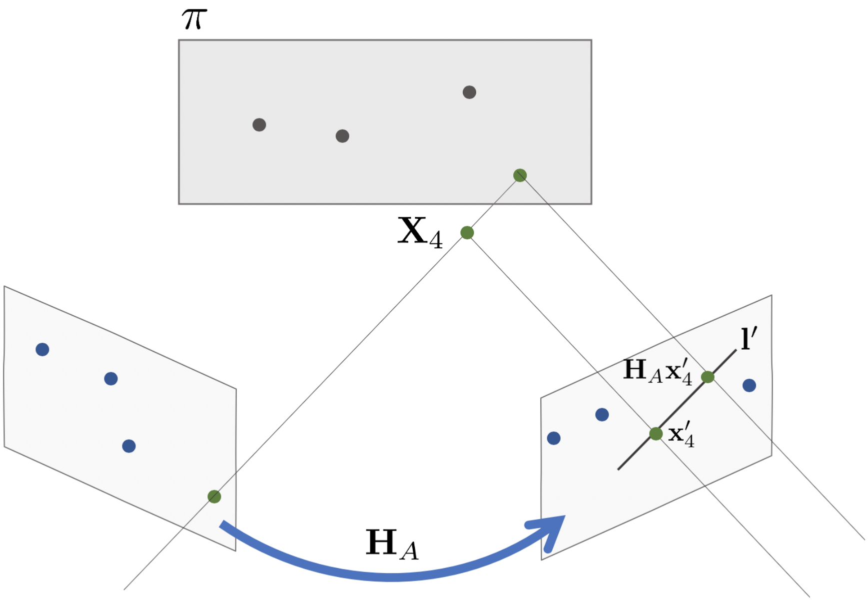
Epipolar lines

Given a point $\mathbf{x}$ on the first Affine image plane and the Affine Fundamental Matrix $\mathbf{F}_A$, the Epipolar Line $\mathbf{l}'$ on the second Affine image plane can be calculated as follows:

$$\begin{aligned} \mathbf{x} &= \begin{pmatrix} x & y & 1 \end{pmatrix}^\intercal \\ \mathbf{F}_A &= \begin{bmatrix} 0 & 0 & a \\ 0 & 0 & b \\ c & d & e \end{bmatrix} \\ \mathbf{l}' &= \mathbf{F}_A\mathbf{x} = \begin{pmatrix} a & b & cx + dy + e \end{pmatrix}^\intercal \end{aligned} \tag{396}$$

The first two terms (a, b) of the Epipolar Line $\mathbf{l}'$ are independent of the point $\mathbf{x} = (x, y, 1)^\intercal$ on the image plane, indicating that the Epipolar Lines are parallel regardless of $\mathbf{x}$.

Estimating $\mathbf{F}_A$ from image point correspondences



To calculate the Fundamental Matrix $\mathbf{F}$ for a Projective camera, at least 8 pairs of corresponding points $(\mathbf{x}, \mathbf{x}')$ between two image planes are required. However, for an Affine camera's Affine Fundamental Matrix $\mathbf{F}_A$, since the top-left 2×2 terms are zero as explained earlier, $\mathbf{F}_A$ can be estimated with only at least 4 pairs of corresponding points $(\mathbf{x}, \mathbf{x}')$.

Algorithm 14.2

- **Objective:** Given 4 pairs of corresponding points $(\mathbf{x}_i, \mathbf{x}'_i),\ i = 1, \cdots, 4$ on the Affine image plane, use these to calculate the Affine Fundamental Matrix.

- Initially, the first three pairs of corresponding points $(\mathbf{x}_i, \mathbf{x}'_i),\ i = 1, 2, 3$ can determine a unique plane π by spanning the world points $\mathbf{X}_i,\ i = 1, 2, 3$ obtained by back-projection.

$$\pi = \text{Span}\{\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3\} \tag{397}$$

- **Calculate the Affine Homography $\mathbf{H}_A$ that satisfies $\mathbf{x}' = \mathbf{H}_A\mathbf{x}$ for the world plane π.** Normally, calculating a Homography requires more than four pairs of corresponding points, but an Affine Homography can be calculated with only three pairs since its last row is always $(0, 0, 1)$.

- Use the remaining point $\mathbf{x}_4$ to calculate the **Epipolar Line $\mathbf{l}'$** connecting $\mathbf{H}_A\mathbf{x}_4$ and $\mathbf{x}'_4$.

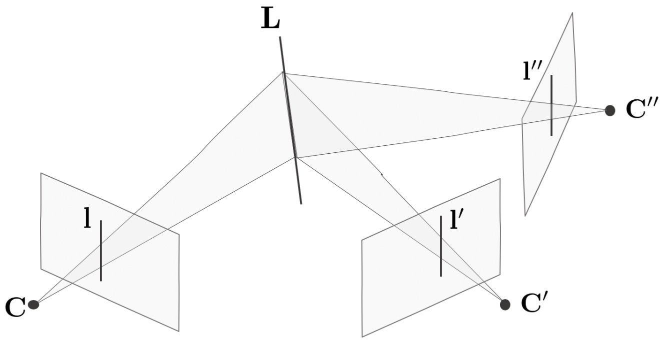
$$\mathbf{l}' = \mathbf{H}_A\mathbf{x}_4 \times \mathbf{x}'_4 \tag{398}$$

  Once the Epipolar Line $\mathbf{l}'$ is calculated, the Epipole $\mathbf{e}' = (-l'_2, l'_1, 0)^\intercal$ can be determined. If the world point $\mathbf{X}_4$ back-projected from $\mathbf{x}_4$ exists on the world plane π, an Epipolar Line cannot be determined.

- $\mathbf{F}_A = \mathbf{e}'^\wedge\mathbf{H}_A$ is used to calculate the Affine Fundamental Matrix.

$$\mathbf{F}_A = \begin{bmatrix} -l'_2 & l'_1 & 0 \end{bmatrix}^{\intercal\wedge} \mathbf{H}_A \tag{399}$$

12 The Trifocal Tensor



Trifocal Tensor $\mathcal{T}$

$$\begin{aligned} \mathbf{l}^\intercal &= \mathbf{l}'^\intercal[\mathbf{T}_1,\ \mathbf{T}_2,\ \mathbf{T}_3]\mathbf{l}'' \\ &= \mathbf{l}'^\intercal\mathcal{T}\mathbf{l}'' \end{aligned}$$

The geometric basis for the trifocal tensor

In Three-View Geometry, the Trifocal Tensor $\mathcal{T}$ plays a role similar to the Fundamental Matrix $\mathbf{F}$ in Two-View Geometry. Like $\mathbf{F}$, $\mathcal{T}$ constrains the three different camera image planes with specific constraints.

Incidence relations for lines

Let's assume a line $\mathbf{L}$ on the world plane and three different cameras observing it. Let's denote the image planes of the three cameras as $\pi_{\mathbf{P}}, \pi_{\mathbf{P}'}, \pi_{\mathbf{P}''}$ respectively, and the projections of $\mathbf{L}$ onto these image planes as lines $\mathbf{l}, \mathbf{l}', \mathbf{l}''$. Let's explore the relationship between $\mathbf{l} \leftrightarrow \mathbf{l}' \leftrightarrow \mathbf{l}''$.

Given the camera matrices $(\mathbf{P}, \mathbf{P}', \mathbf{P}'')$, they can be represented in the Canonical Form as:

$$\begin{aligned} \mathbf{P} &= [\mathbf{I} \mid 0] \\ \mathbf{P}' &= [\mathbf{A} \mid \mathbf{a}_4] \\ \mathbf{P}'' &= [\mathbf{B} \mid \mathbf{b}_4] \end{aligned} \tag{400}$$

including projective ambiguity. The line $\mathbf{L}$ in the world is the intersection line of the planes π, π', π'' back-projected from $\mathbf{l}, \mathbf{l}', \mathbf{l}''$, so if we back-project these lines:

$$\begin{aligned} \pi &= \mathbf{P}^\intercal\mathbf{l} = \begin{pmatrix} \mathbf{l} \\ 0 \end{pmatrix} \\ \pi' &= \mathbf{P}'^\intercal\mathbf{l}' = \begin{pmatrix} \mathbf{A}^\intercal\mathbf{l}' \\ \mathbf{a}_4^\intercal\mathbf{l}' \end{pmatrix} \\ \pi'' &= \mathbf{P}'^\intercal\mathbf{l}'' = \begin{pmatrix} \mathbf{B}^\intercal\mathbf{l}'' \\ \mathbf{b}_4^\intercal\mathbf{l}'' \end{pmatrix} \end{aligned} \tag{401}$$

we can obtain the normal vectors (normal vector) of these planes. If a matrix $\mathbf{M} \in \mathbb{R}^{4\times 3}$ containing these three plane's normal vectors as columns is given as

$$\mathbf{M} = \begin{bmatrix} \pi & \pi' & \pi'' \end{bmatrix} \tag{402}$$

then the column space of $\mathbf{M}$ represents a 2-dimensional plane perpendicular to the world line $\mathbf{L}$, thus the rank of $\mathbf{M}$ is 2. Therefore,

$$\mathbf{M} = \begin{bmatrix} \mathbf{m}_1 & \mathbf{m}_2 & \mathbf{m}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{l} & \mathbf{A}^\intercal\mathbf{l}' & \mathbf{B}^\intercal\mathbf{l}'' \\ 0 & \mathbf{a}_4^\intercal\mathbf{l}' & \mathbf{b}_4^\intercal\mathbf{l}'' \end{bmatrix} \tag{403}$$

where $\mathbf{m}_1 = \alpha\mathbf{m}_2 + \beta\mathbf{m}_3$. Since the first column $(2,1)$ value of the second row of matrix $\mathbf{M}$ is 0, α, β can be derived.

$$0 = k(\mathbf{b}_4^\intercal\mathbf{l}'')\mathbf{m}_2 - (k\mathbf{a}_4^\intercal\mathbf{l}')\mathbf{m}_3 \tag{404}$$

Thus, for any constant k, $\alpha = k(\mathbf{b}_4^\intercal\mathbf{l}'), \beta = k(\mathbf{a}_4^\intercal\mathbf{l}')$. Next, if we expand the first row of the matrix $\mathbf{M}$,

$$\begin{aligned} \mathbf{l} &= (\mathbf{b}_4^\intercal\mathbf{l}'')\mathbf{A}^\intercal\mathbf{l}' - (\mathbf{a}_4^\intercal\mathbf{l}')\mathbf{B}^\intercal\mathbf{l}'' \\ &= (\mathbf{l}''^\intercal\mathbf{b}_4)\mathbf{A}^\intercal\mathbf{l}' - (\mathbf{l}'^\intercal\mathbf{a}_4)\mathbf{B}^\intercal\mathbf{l}'' \end{aligned} \tag{405}$$

If the scalar values $\mathbf{a}_4^\intercal \mathbf{l}', \mathbf{b}_4^\intercal \mathbf{l}''$ are the same when transposed, the ith coordinate value of line $\mathbf{l}$ can be represented as

$$
\begin{aligned}
l_i &= \mathbf{l}'^\intercal(\mathbf{a}_i\mathbf{b}_4^\intercal)\mathbf{l}'' - \mathbf{l}'^\intercal(\mathbf{a}_4\mathbf{b}_i^\intercal)\mathbf{l}'' \\
&= \mathbf{l}'^\intercal(\mathbf{a}_i\mathbf{b}_4^\intercal - \mathbf{a}_4\mathbf{b}_i^\intercal)\mathbf{l}''
\end{aligned} \tag{406}
$$

Replacing $\mathbf{a}_i\mathbf{b}_4^\intercal - \mathbf{a}_4\mathbf{b}_i^\intercal$ with $\mathbf{T}_i$, it can be succinctly expressed as

$$
l_i = \mathbf{l}'^\intercal \mathbf{T}_i \mathbf{l}'' \tag{407}
$$

Definition 15.1

Here, **the set of matrices $\{\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3\}$ represents the matrix representation of the Trifocal Tensor $\mathcal{T}$. Using this, the line l can be re-expressed as follows.**

$$
\begin{aligned}
\mathbf{l}^\intercal &= \mathbf{l}'^\intercal \begin{bmatrix} \mathbf{T}_1 & \mathbf{T}_2 & \mathbf{T}_3 \end{bmatrix} \mathbf{l}'' \\
&= (\mathbf{l}'^\intercal \mathbf{T}_1 \mathbf{l}'', \mathbf{l}'^\intercal \mathbf{T}_2 \mathbf{l}'', \mathbf{l}'^\intercal \mathbf{T}_3 \mathbf{l}'')
\end{aligned} \tag{408}
$$

Homographies induced by a plane

Result 15.2

Let's designate the image planes of three different cameras as $\pi_{\mathbf{P}}, \pi_{\mathbf{P}'}, \pi_{\mathbf{P}''}$. If the line $\mathbf{l}'$ on the second camera's image plane is back-projected to obtain a world plane π', then there exists a Homography $\mathbf{H}_{13}$ that transforms from $\pi_{\mathbf{P}}$ to $\pi_{\mathbf{P}''}$. This section describes how to describe $\mathbf{H}_{13}$ through $\mathbf{l}'$.

The line $\mathbf{l}$ on the first image plane can be expressed as

$$
\mathbf{l}^\intercal = \mathbf{l}'^\intercal \begin{bmatrix} \mathbf{T}_1 & \mathbf{T}_2 & \mathbf{T}_3 \end{bmatrix} \mathbf{l}'' \tag{409}
$$

and simplifying this expression gives $\mathbf{l} = \mathbf{H}_{13}^\intercal \mathbf{l}''$ form

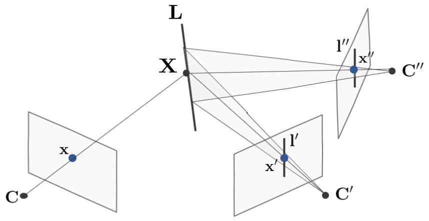
$$
\begin{aligned}
\mathbf{l}^\intercal &= \mathbf{l}'^\intercal \begin{bmatrix} \mathbf{T}_1 & \mathbf{T}_2 & \mathbf{T}_3 \end{bmatrix} \mathbf{l}'' \\
&= (\begin{bmatrix} \mathbf{T}_1^\intercal & \mathbf{T}_2^\intercal & \mathbf{T}_3^\intercal \end{bmatrix} \mathbf{l}')^\intercal \mathbf{l}'' \\
&= \mathbf{H}_{13}^\intercal \mathbf{l}''
\end{aligned} \tag{410}
$$

Thus, $\mathbf{H}_{13} = \begin{bmatrix} \mathbf{T}_1^\intercal, & \mathbf{T}_2^\intercal & \mathbf{T}_3^\intercal \end{bmatrix} \mathbf{l}'$. Using $\mathbf{H}_{13}$, we can perform a Homography transformation from $\pi_{\mathbf{P}} \rightarrow \pi_{\mathbf{P}''}$ such that $\mathbf{x}'' = \mathbf{H}_{13}\mathbf{x}$.

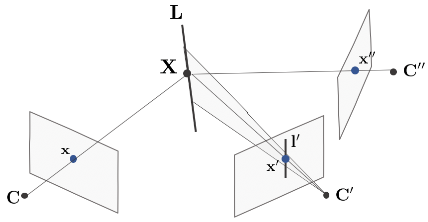
Similarly, $\mathbf{H}_{12}$ satisfies the formula $\mathbf{x}' = \mathbf{H}_{12}\mathbf{x}$ as follows.

$$
\mathbf{H}_{12} = \begin{bmatrix} \mathbf{T}_1 & \mathbf{T}_2 & \mathbf{T}_3 \end{bmatrix} \mathbf{l}'' \quad {}^\forall \mathbf{l}'' \tag{411}
$$

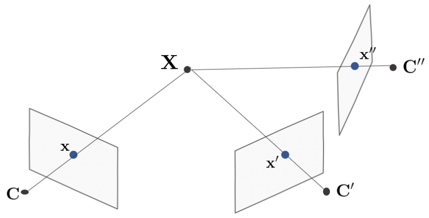
Point and line incidence relations



Point-Line-Line



Point-Line-Point



Point-Point-Point

In the previous section, it was explained that lines existing in three image planes are constrained by the Trifocal Tensor $\mathcal{T}$. **In this section, it is described how not only the three lines but also the relations between points and lines are constrained by the Trifocal Tensor $\mathcal{T}$.**

For any point $\mathbf{x}$ existing on line $\mathbf{l}$, $\mathbf{x}^\intercal\mathbf{l} = 0$ holds and re-expressing it using Tensor notation, we get

$$\begin{aligned} \mathbf{x}^\intercal\mathbf{l} &= \sum_i x^i l_i \\ &= \sum_i x^i \mathbf{l}'^\intercal \mathbf{T}_i \mathbf{l}'' \quad \because l_i = \mathbf{l}'^\intercal \mathbf{T}_i \mathbf{l}'' \\ &= \mathbf{l}'^\intercal (\sum_i x^i \mathbf{T}_i)\mathbf{l}'' = 0 \end{aligned} \tag{412}$$

This can be represented together and it signifies the relationship between **a point and two lines (point-line-line).**

The previously described $\mathbf{H}_{13}$ represents the Homography transforming from the first image plane to the third image plane. This allows us to find the point $\mathbf{x}''$ on the third image plane as

$$\mathbf{x}'' = \mathbf{H}_{13}\mathbf{x} = \begin{bmatrix} \mathbf{T}_1^\intercal\mathbf{l}' & \mathbf{T}_2^\intercal\mathbf{l}' & \mathbf{T}_3^\intercal\mathbf{l}' \end{bmatrix} \mathbf{x} = (\sum_i x^i \mathbf{T}_i^\intercal)\mathbf{l}' \tag{413}$$

which can be expanded as such. Since $\mathbf{x}''$ contains a Scale Factor, by multiplying $\mathbf{x}''^\wedge$ to uniquely determine $\mathbf{x}''$,

$$\mathbf{x}'^\intercal\mathbf{x}''^\wedge = \mathbf{l}'^\intercal (\sum_i x^i \mathbf{T}_i)\mathbf{x}''^\wedge = \mathbf{0}^\intercal \tag{414}$$

occurs. This signifies the relationship between **two points and one line (point-line-point).**
Similarly, when three points are given,

$$\mathbf{x}'^\wedge (\sum_i x^i \mathbf{T}_i)\mathbf{x}''^\wedge = \mathbf{0} \tag{415}$$

allows us to eliminate the Scale Factor and find a unique $\mathbf{x}'$, signifying **the relationship between three points (point-point-point).**

Epipolar lines

Result 15.3

The Trifocal Tensor $\mathcal{T}$ can be used to determine the Epipolar Line of two cameras. Given three image planes $\pi_{\mathbf{P}}, \pi_{\mathbf{P}'}, \pi_{\mathbf{P}''}$ and assuming a point $\mathbf{x}$ exists on $\pi_{\mathbf{P}}$ with Epipolar Lines $\mathbf{l}', \mathbf{l}''$ on planes $\pi_{\mathbf{P}'}, \pi_{\mathbf{P}''}$, the following relationships are established:

$$
\begin{aligned}
\mathbf{l}'^{\intercal}(\sum_i x^i \mathbf{T}_i) = 0 \\
(\sum_i x^i \mathbf{T}_i)\mathbf{l}'' = 0
\end{aligned} \tag{416}
$$

Thus, **$\mathbf{l}'^{\intercal}$ is the Left Null Vector of $(\sum_i x^i \mathbf{T}_i)$ and $\mathbf{l}''$ is the Right Null Vector.**

Proof

If the Epipolar Line $\mathbf{l}'$ on the second image plane $\pi_{\mathbf{P}'}$ is back-projected to create plane π', an intersection line is created between π' and $\pi_{\mathbf{P}}$, becoming the Epipolar Line $\mathbf{l}$ of the $\pi_{\mathbf{P}}$ plane. When any line $\mathbf{l}''$ on the third image plane $\pi_{\mathbf{P}''}$ is back-projected to plane π'', and π', they intersect at line $\mathbf{L}$ on the world plane, and when projected back onto $\pi_{\mathbf{P}}$, $\mathbf{L}$ always projects to a point $\mathbf{x} \in \mathbf{l}$ on the Epipolar Line $\mathbf{l}$.

Through this $\mathbf{l} \leftrightarrow \mathbf{l}' \leftrightarrow \mathbf{l}''$ relationship, the Trifocal Tensor $\mathcal{T}$ can be obtained and the relationship formula between **a point x and two lines (point-line-line) is established.**

$$
\begin{aligned}
& \mathbf{x} \in \mathbf{l} = \mathbf{l}'^{\intercal}\mathcal{T}\mathbf{l}'' \\
& \mathbf{l}'^{\intercal}(\sum_i x^i \mathbf{T}_i)\mathbf{l}'' = 0 \quad {}^{\forall}\mathbf{l}'' \\
& \therefore \mathbf{l}'^{\intercal}(\sum_i x^i \mathbf{T}_i) = 0
\end{aligned} \tag{417}
$$

Since this formula must be satisfied for all $\mathbf{l}''$, the formula $\mathbf{l}'^{\intercal}(\sum_i x^i \mathbf{T}_i) = 0$ holds. Similarly, for all $\mathbf{l}'$, $(\sum_i x^i \mathbf{T}_i)\mathbf{l}'' = 0$ also holds.

Result 15.4

Additionally, the Epipoles $\mathbf{e}', \mathbf{e}''$ can be determined by calculating the intersection of all $\mathbf{l}', \mathbf{l}''$ for any ${}^{\forall}\mathbf{x}$.

Extracting the fundamental matrices

As previously described, the Trifocal Tensor $\mathcal{T}$ allows us to determine the Fundamental Matrix $\mathbf{F}$ for three image planes. This section explains how to derive the Fundamental Matrix $\mathbf{F}$ for three cameras using $\mathcal{T}$.

The Fundamental Matrix $\mathbf{F}_{21}$ can be derived as $\mathbf{F}_{21} = \mathbf{e}'^{\wedge}\mathbf{H}_{21}$. $\mathbf{F}_{ij}$ represents the Fundamental Matrix between the i-th and j-th image planes. $\mathbf{e}'$ is derived as previously described by calculating the Left Null Vector of $\mathbf{T}_i$, obtaining $\mathbf{l}', \mathbf{e}'$, and $\mathbf{H}_{21}$ can be derived as $\begin{bmatrix}\mathbf{T}_1 & \mathbf{T}_2 & \mathbf{T}_3\end{bmatrix}\mathbf{l}''$, therefore

$$
\begin{aligned}
\mathbf{F}_{21} &= \mathbf{e}'^{\wedge}\mathbf{H} \\
&= \mathbf{e}'^{\wedge}\begin{bmatrix}\mathbf{T}_1 & \mathbf{T}_2 & \mathbf{T}_3\end{bmatrix}\mathbf{l}'' \quad {}^{\exists}\mathbf{l}''
\end{aligned} \tag{418}
$$

In this case, $\mathbf{l}''$ must not be in the Null Space of $\mathbf{T}_i$. That is, $\mathbf{T}_i\mathbf{l}'' \neq 0$ must hold. In other words, the rank of $\begin{bmatrix}\mathbf{T}_1 & \mathbf{T}_2 & \mathbf{T}_3\end{bmatrix}\mathbf{l}''$ must be 3. In this case, the Epipole $\mathbf{e}''$ exists in the Null Space of $\mathbf{l}''$, therefore $\mathbf{e}'^{\intercal}\mathbf{l}'' = 0$ is satisfied. Thus,

$$
\mathbf{e}'^{\intercal} \perp \text{Nul } \mathbf{T}_i \quad {}^{\forall}i \tag{419}
$$

holds, which means by substituting $\mathbf{e}'^{\intercal}$ for $\mathbf{l}'^{\intercal}$ always results in the matrix's rank being 3. Consequently, the Fundamental Matrix $\mathbf{F}_{21}$ is

$$
\mathbf{F}_{21} = \mathbf{e}'^{\wedge}\begin{bmatrix}\mathbf{T}_1 & \mathbf{T}_2 & \mathbf{T}_3\end{bmatrix}\mathbf{e}'' \tag{420}
$$

This can be determined through. Similarly, $\mathbf{F}_{31} = \mathbf{e}'^{\intercal\wedge}\begin{bmatrix}\mathbf{T}_1^{\intercal} & \mathbf{T}_2^{\intercal} & \mathbf{T}_3^{\intercal}\end{bmatrix}\mathbf{e}'$ can be derived using the same method.

Retrieving the camera matrices

In Two-view Geometry, given the Fundamental Matrix $\mathbf{F}$, the corresponding pair of camera matrices $(\mathbf{P}, \mathbf{P}')$ can be derived up to projectivity.

$$\begin{aligned}&\mathbf{P} = [\mathbf{I} \mid 0] \\ &\mathbf{P}' = [\mathbf{A} \mid \mathbf{e}'] \\ &\text{where, } \mathbf{F} = \mathbf{e}'^{\wedge}\mathbf{A} \quad \text{in two-view.}\end{aligned} \tag{421}$$

In Three-view Geometry, using the Trifocal Tensor $\mathcal{T}$, $\mathbf{F}_{21} = \mathbf{e}'^{\wedge}\begin{bmatrix}\mathbf{T}_1 & \mathbf{T}_2 & \mathbf{T}_3\end{bmatrix}\mathbf{e}'^{\intercal}$ and $\mathbf{F}_{31} = \mathbf{e}'^{\intercal\wedge}\begin{bmatrix}\mathbf{T}_1^{\intercal} & \mathbf{T}_2^{\intercal} & \mathbf{T}_3^{\intercal}\end{bmatrix}\mathbf{e}'$ have been derived,

$$\begin{aligned}&\mathbf{P} = [\mathbf{I} \mid 0] \\ &\mathbf{P}' = [\begin{bmatrix}\mathbf{T}_1 & \mathbf{T}_2 & \mathbf{T}_3\end{bmatrix}\mathbf{e}'^{\intercal} \mid \mathbf{e}'] \\ &\text{but, } \mathbf{P}'' \neq [\begin{bmatrix}\mathbf{T}_1^{\intercal} & \mathbf{T}_2^{\intercal} & \mathbf{T}_3^{\intercal}\end{bmatrix}\mathbf{e}' \mid \mathbf{e}'^{\intercal}] \quad \text{in three-view.}\end{aligned} \tag{422}$$

This relationship holds, meaning **once $(\mathbf{P}, \mathbf{P}')$ is computed, the world coordinate system becomes fixed, and thus $\mathbf{P}''$ must be re-expressed relative to the fixed world coordinate system.** According to formula (9.10) on p256 of the Canonical Form, the most general camera matrix can be represented as follows:

$$\mathbf{P}'' = [\mathbf{H} + \mathbf{e}'^{\intercal}\mathbf{v}^{\intercal} \mid \lambda\mathbf{e}'^{\intercal}] \tag{423}$$

Expanding this results in

$$\begin{aligned}\mathbf{P}'' &= [\mathbf{H} + \mathbf{e}'^{\intercal}\mathbf{v}^{\intercal} \mid \lambda\mathbf{e}'^{\intercal}] \\ &= [\underbrace{\begin{bmatrix}\mathbf{T}_1^{\intercal} & \mathbf{T}_2^{\intercal} & \mathbf{T}_3^{\intercal}\end{bmatrix}\mathbf{e}' + \mathbf{e}'^{\intercal}\mathbf{v}^{\intercal}}_{\mathbf{B}} \mid \underbrace{\lambda\mathbf{e}'^{\intercal}}_{\mathbf{b_4}}]\end{aligned} \tag{424}$$

And $\mathbf{P}' = [\mathbf{A} \mid \mathbf{a}_4]$ is

$$\begin{aligned}\mathbf{P}' &= [\mathbf{A} \mid \mathbf{a}_4] \\ &= [\underbrace{\begin{bmatrix}\mathbf{T}_1 & \mathbf{T}_2 & \mathbf{T}_3\end{bmatrix}\mathbf{e}'^{\intercal}}_{\mathbf{A}} \mid \underbrace{\mathbf{e}'}_{\mathbf{a}_4}]\end{aligned} \tag{425}$$

Therefore, by the definition of $\mathbf{T}_i$,

$$\begin{aligned}\mathbf{T}_i &= \mathbf{a}_i\mathbf{b}_4^{\intercal} + \mathbf{a}_4\mathbf{b}_i^{\intercal} \\ &= \mathbf{T}_i\mathbf{e}'^{\intercal}\mathbf{e}'^{\intercal\intercal} - \mathbf{e}'\mathbf{b}_i^{\intercal}\end{aligned} \tag{426}$$

This concludes. When simplified,

$$\mathbf{T}_i(\mathbf{I} - \mathbf{e}'^{\intercal}\mathbf{e}'^{\intercal\intercal}) = -\mathbf{e}'\mathbf{b}_i^{\intercal} \tag{427}$$

This occurs. Assuming $\|\mathbf{e}'\| = 1$ and multiplying both sides by $\mathbf{e}'^{\intercal}$ results in

$$\mathbf{b}_i^{\intercal} = \mathbf{e}'^{\intercal}\mathbf{T}_i(\mathbf{e}'^{\intercal\intercal} - \mathbf{I}) \tag{428}$$

Ultimately, $\mathbf{P}''$ is as follows.

$$\mathbf{P}'' = [(\mathbf{e}'^{\intercal}\mathbf{e}'^{\intercal\intercal} - \mathbf{I})\begin{bmatrix}\mathbf{T}_1^{\intercal} & \mathbf{T}_2^{\intercal} & \mathbf{T}_3^{\intercal}\end{bmatrix}\mathbf{e}' \mid \mathbf{e}'^{\intercal}] \tag{429}$$

The trifocal tensor and tensor notation

The trifocal tensor $\mathcal{T}$ can be represented using tensor notation as follows:

$$\begin{aligned}\mathcal{T}_i^{jk} &= (j,k) \text{ entry of } \mathbf{T}_i \\ &= a_i^j b_4^k - a_4^j b_i^k\end{aligned} \tag{430}$$

The i-th coordinate of the line $\mathbf{l}$ represented as a matrix is $l_i = \mathbf{l}'^{\intercal}\mathbf{T}_i\mathbf{l}''$ and can be represented using tensor notation as:

$$
\begin{aligned}
l_i &= l'_i \mathcal{T}_i^{jk} l''_k \\
&= l'_i l''_k \mathcal{T}_i^{jk}
\end{aligned} \tag{431}
$$

Using tensor notation, we can derive a Homography $\mathbf{H} : \pi_{\mathbf{P}} \mapsto \pi_{\mathbf{P}''}$ that transforms the line $\mathbf{l}$ from the first camera to the third camera:

$$
l_i = l''_k (l'_j \mathcal{T}_i^{jk}) = l''_k h_i^k \tag{432}
$$

Here, $h_i^k = l'_j \mathcal{T}_i^{jk}$. When using Homography $\mathbf{H}$ to transform a point:

$$
x''^k = h_i^k x^i \tag{433}
$$

An explanation of ϵ_{ijk} useful in tensor representation is:

$$
\epsilon_{ijk} = \begin{cases} 0 & \text{unless i,j,k are all distinct} \\ +1 & \text{if i,j,k are an even permutation of 1,2,3} \\ -1 & \text{if i,j,k are an odd permutation of 1,2,3} \end{cases} \tag{434}
$$

That is, ϵ_{ijk} is 0 unless i, j, k are all distinct, and takes a value of $+1$ when i, j, k are in a sequence such as $(1, 2, 3), (3, 1, 2)$ or $(2, 3, 1)$, and -1 otherwise. The cross product of 3×3 vectors can be expressed as $\mathbf{c} = \mathbf{a} \times \mathbf{b}$:

$$
c_i = \epsilon_{ijk} a^j b^k \tag{435}
$$

This represents $\begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} a_2 b_3 - a_3 b_2 \\ a_3 b_1 - a_1 b_3 \\ a_1 b_2 - a_2 b_1 \end{bmatrix}$ using tensor notation. Thus, $(\mathbf{a}^\wedge)_{ik}$ in tensor notation:

$$
(\mathbf{a}^\wedge)_{ik} = \epsilon_{ijk} a^j \tag{436}
$$

The trilinearities

Using tensor notation, the relationships between various points and lines discussed in the previous section can be re-expressed. For example, the general form of the relationship between two points and one line (point-line-point) can be represented as:

$$
\mathbf{l}'^\intercal (\sum_i x^i \mathbf{T}_i) \mathbf{x}''^\wedge = \mathbf{0}^\intercal \tag{437}
$$

When expressed in tensor notation, $(\mathbf{x}''^\wedge)_{qs} = -x''^k \epsilon_{kqs}$, which reorganized as:

$$
l'_j x^i \mathcal{T}_i^{jq} x''^k \epsilon_{kqs} = 0_s \tag{438}
$$

This is referred to as trilinearities because it derives a linear equation using three different points or lines from three distinct images.

Transfer

When given three cameras, using the Trifocal Tensor $\mathcal{T}$ to determine the position of a point or line on another image plane when the positions on two of the image planes are known is referred to as Transfer.

Point transfer using fundamental matrices

Point Transfer refers to determining the position of $\mathbf{x}''$ using the Fundamental Matrix $\mathbf{F}_{21}, \mathbf{F}_{31}, \mathbf{F}_{32}$ given for the three image planes $\pi_{\mathbf{P}}, \pi_{\mathbf{P}'}, \pi_{\mathbf{P}''}$, based on the known positions of $\mathbf{x}, \mathbf{x}'$. This can be determined through Epipolar Geometry. The Epipolar Line $\mathbf{l}''$ existing on the third image plane $\pi_{\mathbf{P}''}$ is:

$$
\begin{aligned}
\mathbf{l}''_{31} &= \mathbf{F}_{31} \mathbf{x} \\
\mathbf{l}''_{32} &= \mathbf{F}_{32} \mathbf{x}'
\end{aligned} \tag{439}
$$

Therefore, a point $\mathbf{x}''$ on $\pi_{\mathbf{P}''}$ must satisfy:

$$\mathbf{x}'' \in \mathbf{l}''_{31} \text{ and } \mathbf{l}''_{32} \tag{440}$$

Thus, $\mathbf{x}''$ is determined as the intersection of these two lines:

$$\mathbf{x}'' = (\mathbf{F}_{31}\mathbf{x}) \times (\mathbf{F}_{32}\mathbf{x}') \tag{441}$$

Although the formula $\mathbf{F}_{21}$ was not used above, in reality, it is used to improve results as the correspondence pair $(\mathbf{x}, \mathbf{x}')$ may contain noise, preventing the equation $\mathbf{x}^\intercal \mathbf{F}_{21}\mathbf{x}' = 0$ from being satisfied. Therefore, the Optimal Triangulation method discussed in the previous section is used to minimize the value of $d(\mathbf{x}, \mathbf{l}(t))^2 + d(\mathbf{x}', \mathbf{l}'(t))^2$, determining $\hat{\mathbf{x}} \leftrightarrow \hat{\mathbf{x}}'$ and then calculating $\mathbf{x}''$:

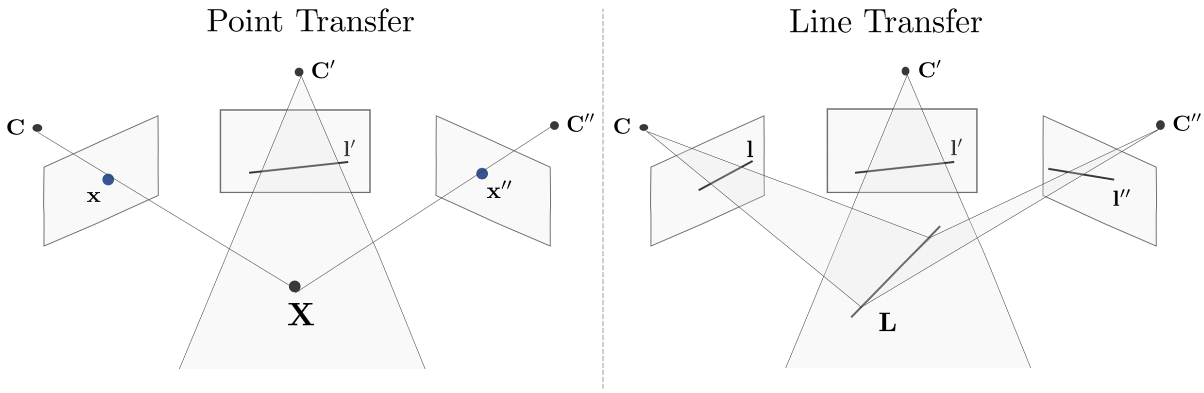
$$\mathbf{x}'' = (\mathbf{F}_{31}\hat{\mathbf{x}}) \times (\mathbf{F}_{32}\hat{\mathbf{x}}') \tag{442}$$

However, **if there exist corresponding point pairs $\mathbf{x}, \mathbf{x}'$ on the Trifocal plane formed by the centers $\mathbf{C}, \mathbf{C}', \mathbf{C}''$ of the three cameras, the Epipolar Line projected onto $\pi_{\mathbf{P}''}$ is generated identically as follows:**

$$\mathbf{F}_{31}\mathbf{x} = \mathbf{F}_{32}\mathbf{x}' \tag{443}$$

This implies that the unique position of $\mathbf{x}''$ cannot be determined, which can be seen as a limitation of Point Transfer using the Fundamental Matrix.

Point transfer using the trifocal tensor



Using the Trifocal Tensor $\mathcal{T}$ instead of the Fundamental Matrix $\mathbf{F}_{ij}$ allows for transfer in a broader range of cases. The method of Point Transfer using $\mathcal{T}$ is as follows:

- Calculate $\mathbf{F}_{21}$ using $\mathcal{T}$.
$$\mathbf{F}_{21} = \mathbf{e}'^\intercal \begin{bmatrix} \mathbf{T}_1 & \mathbf{T}_2 & \mathbf{T}_3 \end{bmatrix} \mathbf{e}'' \tag{444}$$
- Use the Optimal Triangulation method to calculate the optimal corresponding point pairs with reduced noise.
$$(\mathbf{x}, \mathbf{x}') \rightarrow (\hat{\mathbf{x}}, \hat{\mathbf{x}'}) \tag{445}$$
- Calculate the Epipolar Line $\mathbf{l}'_{\mathbf{e}}$ on the second image plane $\pi_{\mathbf{P}'}$. Next, calculate the line $\mathbf{l}'$ that is perpendicular to $\mathbf{l}'_{\mathbf{e}}$ and passes through $\hat{\mathbf{x}}'$.

$$\begin{aligned} \hat{\mathbf{x}}' = (x'_1, x'_2, 1)^\intercal \qquad\qquad & \mathbf{l}'_{\mathbf{e}} = \begin{bmatrix} l_1 & l_2 & l_3 \end{bmatrix}^\intercal \\ \Rightarrow \mathbf{l}' = \begin{bmatrix} l_2 \\ -l_1 \\ -l_2 x'_1 + l_1 x'_2 \end{bmatrix} & \end{aligned} \tag{446}$$

- Use $\mathbf{l}'$ to perform Point Transfer of $\mathbf{x}'$.

$$x''^k = x'^i l'_j \mathcal{T}_i^{jk} \tag{447}$$

Degenerate configurations

Even when using the Trifocal Tensor $\mathcal{T}$, if a point $\mathbf{X}$ in 3D space lies on the baseline connecting the centers $\mathbf{C}, \mathbf{C}'$ of two cameras, the position of $\mathbf{x}''$ cannot be determined.

Line transfer using the trifocal tensor

The Trifocal Tensor $\mathcal{T}$ can be used to transfer not only points but also lines. If the lines $\mathbf{l} \leftrightarrow \mathbf{l}' \leftrightarrow \mathbf{l}''$ on the three image planes correspond as follows:

$$l_i = l'_j l''_k \mathcal{T}_i^{jk} \tag{448}$$

this can be expressed through $\mathcal{T}$ and implies that the line $\mathbf{l}$ is parallel to the vector $[l'_j l''_k \mathcal{T}_i^{jk}]$.

$$\mathbf{l} \ \parallel \ [l'_j l''_k \mathcal{T}_i^{jk}]_{3\times 1}, \ i = 1, 2, 3 \tag{449}$$

Parallel lines imply that the Cross Product must be zero:

$$\begin{aligned} (\mathbf{l}^{\wedge})_{si} l'_j l''_k \mathcal{T}_i^{jk} = 0 \\ (l_s \epsilon^{ris}) l'_j l''_k \mathcal{T}_i^{jk} = 0 \end{aligned} \tag{450}$$

This is a linear equation for $\mathbf{l}''$, and reorganizing gives:

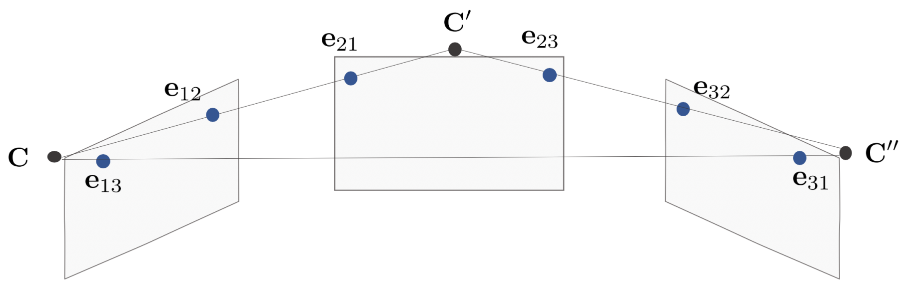
$$(l_s \epsilon^{ris} l'_j \mathcal{T}_i^{jk}) l''_k = 0 \tag{451}$$

By solving the linear equation, $\mathbf{l}''$ can be calculated when $\mathbf{l} \leftrightarrow \mathbf{l}'$ is given.

Degeneracies

If the lines $\mathbf{l}, \mathbf{l}'$ are Epipolar Lines, the back-projected plane π will be the same Epipolar Plane as π', making it impossible to uniquely determine the line $\mathbf{l}''$ on the third image plane $\pi_{\mathbf{P}''}$.

The fundamental matrices for three views



When the Fundamental Matrices $\mathbf{F}_{21}, \mathbf{F}_{31}, \mathbf{F}_{32}$ for three cameras are given, they are not independent of each other. Six Epipoles are created for the three image planes:

$$\mathbf{e}_{23}^{\intercal} \mathbf{F}_{21} \mathbf{e}_{13} = \mathbf{e}_{31}^{\intercal} \mathbf{F}_{32} \mathbf{e}_{21} = \mathbf{e}_{32}^{\intercal} \mathbf{F}_{31} \mathbf{e}_{12} = 0 \tag{452}$$

Definition 15.5

For the three Fundamental Matrices $\mathbf{F}_{21}, \mathbf{F}_{31}, \mathbf{F}_{32}$ to be independent and compatible, the above formula must hold.

Uniqueness of camera matrices given three fundamental matrices

If the three Fundamental Matrices are compatible, the three camera matrices $(\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3)$ that generate them are determined uniquely up to projectivity:

$$\begin{aligned} &{}^{\exists}(\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3) \ \text{ such that the fundamental matrix of } (\mathbf{P}_i, \mathbf{P}_j) \ \text{ is } \ \mathbf{F}_{ij} \\ &(\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3) \ \text{ are unique up to projectivity.} \end{aligned} \tag{453}$$

To prove this, proceed as follows:

- Using the principles of Two-view Geometry, when $\mathbf{F}_{21} = \mathbf{e}_{21}^{\wedge}\mathbf{A}$, camera matrices $\mathbf{P}_1, \mathbf{P}_2$ can be determined as follows:

$$\begin{aligned} \mathbf{P}_1 &= [\mathbf{I} \mid 0] \\ \mathbf{P}_2 &= [\mathbf{A} \mid \mathbf{e}_{21}] \end{aligned} \tag{454}$$

- Generate corresponding point pairs $(\mathbf{x}_i, \mathbf{x}'_i)$ that satisfy $\mathbf{x}'^{\intercal}\mathbf{F}_{21}\mathbf{x} = 0$. Triangulate the world space point $\mathbf{X}_i$ through this.
- Use the point-point-point relationship formula to determine $\mathbf{x}''_i$.

$$\mathbf{x}''_i = (\mathbf{F}_{31}\mathbf{x}_i) \times (\mathbf{F}_{32}\mathbf{x}'_i) \tag{455}$$

- Use the formula $\mathbf{P}_3\mathbf{X}_i = \mathbf{x}''_i$ to calculate $\mathbf{P}_3$.

However, if the world space point $\mathbf{X}_i$ exists on the Trifocal plane including the centers $\mathbf{C}, \mathbf{C}', \mathbf{C}''$ of the three cameras, the unique position of $\mathbf{x}''$ cannot be determined.

13 Revision log

- 1st: 2020-05-12
- 2nd: 2020-06-06
- 3rd: 2020-06-07
- 4th: 2020-06-09
- 5th: 2020-06-10
- 6th: 2020-06-11
- 7th: 2020-06-12
- 8th: 2020-06-14
- 9th: 2020-06-15
- 10th: 2020-06-16
- 11th: 2020-06-20
- 12th: 2020-06-22
- 13th: 2020-06-23
- 14th: 2022-06-28
- 15th: 2022-12-20
- 16th: 2023-01-01
- 17th: 2023-01-21
- 18th: 2024-03-20
- 19th: 2024-03-23
- 20th: 2024-05-04

14 References

Hartley, Richard, and Andrew Zisserman. Multiple view geometry in computer vision. Cambridge university press, 2003

15 Closure

Check out alida.tistory.com for the web version posting.