
NOTES ON 3D GEOMETRY FOR COMPUTER VISION AND SLAM

CONTENTS

PROJECTIVE GEOMETRY
IMAGE RECTIFICATION
SLAM

BY

0xLeo (github.com/0xleo)

AUGUST 14, 2021

DRAFT X.YY
MISSING: ...

Contents

1 Projective geometry	2
1.1 Homogeneous coordinates	2
1.1.1 Motivation	2
1.1.2 Definition	3
1.1.3 Converting between Euclidean and homogeneous coordinates	3
1.1.4 Geometric interpretation of homogeneous points	4
1.2 Mathematics of lines and points in 3D	4
1.2.1 Geometric interpretation of conversions between E.C. and H.C.	7
1.3 The cross ratio	8
1.3.1 The maths of cross ratio	8
1.3.2 Cross ratio and single view metrology	11
1.3.3 Cross ratio and height forensics	12
A Appendices	15
A.1 Pinhole camera model derivation	16

1 Projective geometry

1.1 Homogeneous coordinates

1.1.1 Motivation

Homogeneous coordinates (H.C.) are a way of representing points and lines. In the homogeneous space, it is easier to formalise some mathematical derivations than in the Euclidean space. Homogeneous coordinates are often used in computer vision mainly in two broad applications.

1. They allow us to *compactly represent point transforms*. For instance, they allow us to represent the affine transform (rotation and translation) as a matrix multiplication $\mathbf{y} = \mathbf{Ax}$ as opposed to $\mathbf{y} = \mathbf{Rx} + \mathbf{T}$, where \mathbf{A} (affine matrix) embeds \mathbf{R} (rotation) and \mathbf{T} (translation). To see why this is the case, consider the rotation and translation mapping of a point $[x_1 \ x_2 \ x_3]^\top$ given by:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \mapsto \underbrace{\begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}}_{\mathbf{R}} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}}_{\mathbf{x}} + \underbrace{\begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix}}_{\mathbf{T}}$$

This can be written as a single matrix multiplication, which is more compact and also allows computer programs to optimise it.

The trick is to pack \mathbf{R} and \mathbf{T} into one 3×4 matrix $[\mathbf{R} | \mathbf{T}]$ by appending \mathbf{T} to \mathbf{R} :

$$[\mathbf{R} | \mathbf{T}] = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix}$$

This is the so-called affine matrix \mathbf{A} . However, the multiplication of \mathbf{A} with \mathbf{x} is undefined as \mathbf{x} is 3×1 . The can augment \mathbf{x} to the size 4×1 by introducing a new (redundant) coordinate $x_4 = 1$. But now the product $[\mathbf{R} | \mathbf{T}] \mathbf{x}$ is of size 3×1 , which certainly does not belong in the homogeneous space of the original vector of length 4. So to keep $[\mathbf{R} | \mathbf{T}] \mathbf{x}$ an inner operation and to satisfy $\mathbf{x}_{hom} = \mathbf{I}_{hom} \mathbf{x}_{hom}$ we add a fourth redundant row $[0 \ 0 \ 0 \ 1]$. The final matrix

$$\mathbf{A} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

is an affine matrix. $[x_1 \ x_2 \ x_3 \ 1]$ is the H.C. form the Euclidean point $[x_1 \ x_2 \ x_3]$. Using the affine matrix, translation can also be expressed as part of a matrix product.

$$\mathbf{y} = \mathbf{Ax} = \begin{bmatrix} \mathbf{R} & | & \mathbf{T} \\ 0 & 0 & 0 & 1 \end{bmatrix} \mathbf{x}$$

2. They can *represent points at the projective plane*, i.e. 3D points from the “real world” mapped into a 2D plane (screen). They can even represent points at infinity via finite scalars; without running into divisions by zero. Projections ubiquitously take place in camera models, the simplest one being a pinhole camera.

Camera models, such as the pinhole map a 3D point (x, y, z) in the “real world” onto a 2D plane (camera film/screen) using linear transforms. The linear transform depends on the internal manufacturing characteristics (a.k.a. *intrinsics*) of the camera and the externals (*extrinsics*), such as its translation and rotation. For the pinhole model, the only intrinsic we consider is the focal length f^1 . It can be easily derived (see A.1) that the pinhole camera projection is described by the following transform.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \frac{fx}{z} \\ \frac{fy}{z} \end{bmatrix} = \frac{1}{z} \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad (1.1)$$

H.C. allow us to express Eq. (1.1) as a linear transform, hence more concisely. Although Eq. (1.1) is simple, camera models can get complicated.

¹The projected coordinates are in distance units for simplicity, so that we can stick to one intrinsic – the focal length. If they were measured in pixels, we would include two more intrinsic parameters in the matrix.

Apart from points, H.C. also represent lines (via their coefficients). Lines appear virtually everywhere in images (e.g. edges). Homogeneous coordinates help us answer questions about the correspondences of lines between the 3D world and the image plane, such as;

- Do parallel lines in the real world meet in the projection?
- Do angles between lines in the real world get preserved in the projection?
- Do straight lines stay straight?

Spoiler; the answers are yes, no, and yes (at least for the pinhole model).



Fig. 1. Two properties of the pinhole-based projection; parallel lines that are not parallel to the horizon meet, straight lines remain straight (project to straight lines).

1.1.2 Definition

Homogeneous coordinates are useful for representing an object up to a scale. Mathematically, if \mathbf{x} represents a point in the H.C. space, then $\lambda\mathbf{x}$, $\lambda \neq 0$ must represent the same point, i.e. \mathbf{x} and $\lambda\mathbf{x}$ must be equivalent (\sim) in H.C. We denote equivalencies in H.C. with \sim , therefore $\mathbf{x} \sim \lambda\mathbf{x}$. A point in the H.C. space is defined only by its direction and not by its magnitude. The H.C. for 3D points are hence defined as follows.

DEFINITION 1.1 (Homogeneous Coordinates of a 3D vector). A Euclidean 3D point $\mathbf{x} = (x, y, z)^\top$ is represented in homogeneous coordinates by ANY 4-vector $(x_1, x_2, x_3, x_4)^\top$ such that:

$$x = \frac{x_1}{x_4}, \quad y = \frac{x_2}{x_4}, \quad z = \frac{x_3}{x_4}, \quad \text{in H.C.: } \mathbf{x} = (x_1, x_2, x_3, x_4)^\top \quad (1.2)$$

x_4 can attain any value. For instance for $x_4 = 1$ the H.C. of (x, y, z) are $(x, y, z, 1)$. The following lemma is therefore straightforward.

LEMMA 1.1. In H.C., $[x_1 \ x_2 \ x_3 \ x_4]^\top$ and $\lambda[x_1 \ x_2 \ x_3 \ x_4]^\top$, $\lambda \neq 0$ represent the same point, i.e.

$$\mathbf{x} \sim \lambda\mathbf{x}, \quad \mathbf{x} = [x_1 \ x_2 \ x_3 \ x_4]^\top, \quad \lambda \neq 0 \quad (1.3)$$

For example, $(2, 3, 4, 1) \sim (6, 9, 12, 3) \sim (2, 3, 4)$, and $(4, 8, 12, 2) \sim (2, 4, 6)$.

1.1.3 Converting between Euclidean and homogeneous coordinates

From Lemma 1.1, it is clear that to convert a point (x, y, z) from Euclidean to H.C. we simply add one more ordinate $w = 1$ and obtain $(x, y, z, 1)$. From now on black dot (\bullet) denotes H.C. and white (\circ) E.C.

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \circ \bullet \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ 1 \end{bmatrix} \quad (1.4)$$

Inversely, to “dehomogenise” a point $(x_1, x_2, x_3, x_4)^\top$, $x_4 \neq 0$, we first divide its elements by x_4 such that the last element is 1. We then discard the last element to obtain $(x_1/x_4, x_2/x_4, x_3/x_4)$.

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \bullet \circ \begin{bmatrix} x_1/x_4 \\ x_2/x_4 \\ x_3/x_4 \end{bmatrix}, \quad x_4 \neq 0 \quad (1.5)$$

Upright style \mathbf{x} will be used for H.C. vectors and italic \mathbf{x} for Euclidean.

Converting from E.C. to H.C. is NOT unique! Converting H.C. to E.C. is unique.

For example, $(1, 2, 3, 4)^\top \sim (3, 6, 9, 12)^\top \sim (-2, -4, -6, -8) \bullet \circ (0.25, 0.5, 0.75)$ and they all represent the 3D point. Also $(0, 0, 2) \sim (0, 0, \pi) \bullet \circ (0, 0)$.

• for H.C., \circ for E.C.

EXAMPLE 1.1. Calculate the midpoint between $\mathbf{p}_1 = (5, 7, 8)$ and $\mathbf{p}_2 = (4, -6, 1)$.

SOLUTION 1.1. We convert to E.C. to find the midpoint.

$$\mathbf{p}_1 = (5, 7, 8) \bullet \circ (5/8, 7/8) := \mathbf{p}_1,$$

$$\mathbf{p}_2 = (4, -6, 1) \bullet \circ (4, -6) := \mathbf{p}_2$$

The midpoint in E.C. is $(\mathbf{p}_1 + \mathbf{p}_2)/2 = (37/16, -41/16) \bullet \circ (37/16, -41/16, 1) \sim (37, -41, 16)$

Distance ratios are not preserved in when converting between E.C. and H.C.

1.1.4 Geometric interpretation of homogeneous points

If we were to convert a 2D point in E.C. (x_1, y_1) to H.C., we would augment:

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \circ \bullet \begin{bmatrix} x_1 \\ x_2 \\ 1 \end{bmatrix}$$

Therefore in a 3D plane, the point $(x_1, x_2, 1)$ in H.C. can be geometrically determined by looking at the direction (x_1, x_2) and at height $x_3 = 1$ from the origin. Of course, because $(x_1, x_2, 1) \sim \lambda(x_1, x_2, 1)$, point $(x_1, x_2, 1)$ is also found at $(2x_1, 2x_2, 2)$, $(2.5x_1, 2.5x_2, 2.5)$, etc. but as a convention we will be taking $x_3 = 1$.

Projecting onto $x_3 = 1$ helps us also find the E.C. representation of the projected point – if $x_3 = 1$, then the E.C. are simply (x_1, x_2) . Because the projected point is expressed as $(x/w, y/w, 1)$ (Fig. 2), if we plug in $x = 0, y = 0$, then we observe that the origin $O_2(0, 0)$ in E.C. gets mapped to $O_3(0, 0, 1)$ in H.C.

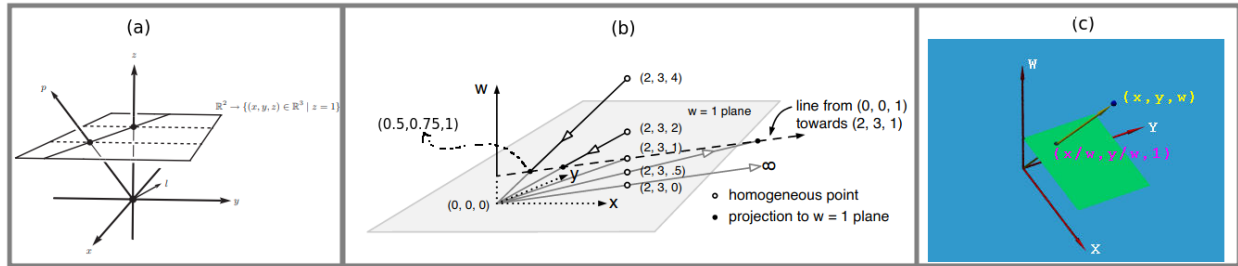


Fig. 2. (a): The set of homogeneous points (x, y, z) in the \mathbb{R}^3 space. (b): Numerical example of homogeneous points projected on $z = 1$. (c): To project to $z = 1$, we divide all ordinates by w .

1.2 Mathematics of lines and points in 3D

We have to describe the mathematics of lines in the projected space (2D) since this is what cameras use. As we will see, H.C. make it easier than Euclidean. Suppose we have a 2D point represented in E.C. as $\mathbf{x} = (x, y)$. If this point lies on a line, then this line is defined by some a_1, a_2, a_3 such that: $a_1x + a_2y + a_3 = 0$, $a_1 \neq 0$ or $a_2 \neq 0$. We can rewrite this equation for in H.C. if we convert $\mathbf{x} = (x, y) \circ \bullet \mathbf{x} = (x_1, x_2, x_3)$ by following Eq. (1.2):

$$x = \frac{x_1}{x_3}, \quad y = \frac{x_2}{x_3}, \quad x_3 \in \mathbb{R}^*$$

Then the line is written in H.C. as:

$$a_1x + a_2y + a_3 = 0 \Rightarrow$$

$$a_1 \frac{x_1}{x_3} + a_2 \frac{x_2}{x_3} + a_3 = 0 \Rightarrow$$

$$a_1x_1 + a_2x_2 + a_3x_3 = 0 \Rightarrow$$

$$\ell^\top \mathbf{x} = 0, \quad \ell = (a_1, a_2, a_3)^\top$$

$s\ell, s \in \mathbb{R}^*$ and ℓ encode the same line.

COROLLARY 1.1 (line equation in H.C.). If a 2D point (x, y) in E.C. lies on a line parametrised by (a_1, a_2, a_3) , then in H.C. the line equation is described by a dot product:

$$\ell^\top \mathbf{x} = 0, \quad \ell^\top = (a_1, a_2, a_3)^\top, \quad a_1, a_2 \neq 0 \quad (1.6)$$

Eq. (1.6) allows us to concisely express the intersection of two lines and the line through two points.

COROLLARY 1.2 (intersection of lines in H.C.). Consider the two parametrised lines $\ell_1 = (a_1, a_2, a_3)^\top$, $\ell_2 = (b_1, b_2, b_3)^\top$ in H.C. Then they intersect at point $\mathbf{p} = \ell_1 \times \ell_2$, where \times denotes cross product.

Proof. Let \mathbf{p} be the point (in H.C.) where ℓ_1 and ℓ_2 intersect. Therefore \mathbf{p} satisfies the line equation Eq. (1.6) for both lines

$$\mathbf{p} \cdot \ell_1 = 0$$

$$\mathbf{p} \cdot \ell_2 = 0$$

, where $\mathbf{p} \cdot \ell_1 = \mathbf{p}^\top \ell_1$ is a dot product. We don't care about the magnitude of \mathbf{p} , we only need such a vector that is perpendicular to both ℓ_1 and ℓ_2 . Such a vector is by its definition the cross product $\ell_1 \times \ell_2$, i.e.

$$\underbrace{(\ell_1 \times \ell_2)}_{\mathbf{p}} \cdot \ell_1 = \underbrace{(\ell_1 \times \ell_2)}_{\mathbf{p}} \cdot \ell_2 = 0$$

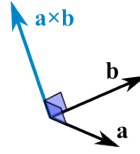


Fig. 3. The cross product $\mathbf{a} \times \mathbf{b}$ is perpendicular to both \mathbf{a} and \mathbf{b} .

□

COROLLARY 1.3 (line joining two points in H.C.). If $\mathbf{p}_1, \mathbf{p}_2$ are two points in H.C., then the coefficients of the line joining them can be computed from $\ell = \mathbf{p}_1 \times \mathbf{p}_2$.

Proof. We want both \mathbf{p}_1 and \mathbf{p}_2 to lie on the same line, i.e. satisfy its equation:

$$\mathbf{p}_1 \cdot \ell = 0$$

$$\mathbf{p}_2 \cdot \ell = 0$$

As shown the previous proof, to satisfy these two conditions we can choose $\ell = \mathbf{p}_1 \times \mathbf{p}_2$.

□

Note that Cor. 1.3 is *dual* to Cor. 1.2, in the sense that points and lines are swapped. This occurs because both points and lines are represented by 3D vectors. Fig. 4 visualises geometrically Cor. 1.3 and Cor. 1.2 in H.C.

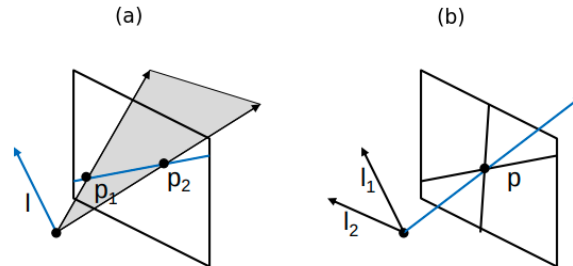


Fig. 4. (a): The vector \mathbf{l} with the coefficients of the line joining $\mathbf{p}_1, \mathbf{p}_2$ is perpendicular to both \mathbf{p}_1 and \mathbf{p}_2 vectors, therefore perpendicular to the plane they span. (b): The same principle holds for a point \mathbf{p} which is at the intersection of lines $\mathbf{l}_1, \mathbf{l}_2$.

Points at infinity ($\pm\infty, \pm\infty$) can be represented in H.C. by finite quantities. To verify that, consider all points starting from $\mathbf{r}_0 = (x_0, y_0)$ with direction $\mathbf{v} = (a, b)$. Then, points given by $\mathbf{r} = \mathbf{r}_0 + t\mathbf{v}$, $t \in \mathbb{R}$ lie along the same straight line. Each scalar component of \mathbf{r} is given by:

$$x = x_0 + ta, \quad y = y_0 + tb \quad (1)$$

Eq. (1) is the parametric line equation of the 2D line where points $\mathbf{r} = (x, y, z) = (x_0 + ta, y_0 + tb, z_0 + tc)$ lie. In H.C., points of \mathbf{r} are given by $\mathbf{r} = (x_0 + ta, y_0 + tb, 1)$. Because $\mathbf{r} = \mathbf{r}/t$ in H.C., they are also given by:

$$\mathbf{r} = \left(\frac{x_0}{t} + a, \frac{y_0}{t} + b, \frac{1}{t} \right) \quad (2)$$

The E.C. points $(x_0 + ta, y_0 + tb, z_0 + tc)$ approach infinity as $t \rightarrow \infty$. If we plug in $t \rightarrow \infty$ in Eq. (2), it turns out the points in infinity in H.C. are given by:

$$\mathbf{r} = (a, b, 0)$$

Moving in the directions of (a, b) and $(-a, -b)$ will end up at the same infinity point represented by $(a, b, 0)$ as well as $(-a, -b, 0)$. For instance, the points at infinity in the direction vector $\mathbf{v} = (-3, 2)$ are given by $(-3, 2, 0)$ in H.C. and all lie on the line $y = -2/3x$.

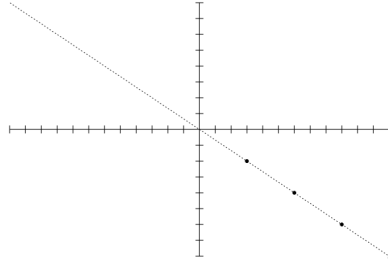


Fig. 5. Points at infinity along the direction $(-3, 2)$

COROLLARY 1.4 (point at infinity in H.C.). In H.C., points of the form $(a, b, 0)$ represent a point that tends to infinity in E.C. in the direction of $\mathbf{v} = (a, b)$. They're also known as *ideal points* (or *vanishing points*).

Ideal points all lie along the same line ℓ . If $\ell = (l_1, l_2, l_3)$ in H.C., then all ideal points $(a, b, 0)$ must satisfy

$$[l_1 \quad l_2 \quad l_3]^\top [a \quad b \quad 0] = 0 \Rightarrow a = 0, \quad b = 0, \quad c \in \mathbb{R}$$

Therefore all ideal points lie along the line with coefficients $(0, 0, c) \sim (0, 0, 1)$ in H.C. To summarise:

COROLLARY 1.5 (line at infinity). All ideal points lie along the line $\ell_\infty = (0, 0, 1)$ in H.C. This is known as *ideal line* (or *vanishing line*).

Ideal points are important because that's where lines that are parallel in E.C. meet. To verify that, consider the equations of two parallel lines:

$$\begin{cases} a_1x + b_1y + c_1 = 0 \\ a_1x + b_1y + c_2 = 0 \end{cases}, \quad c_1 \neq c_2$$

Therefore in H.C., they can be described by $\ell_1 = (a_1, b_1, c_1)$, $\ell_2 = (a_1, b_1, c_2)$. From Cor. 1.2, we know that they intersect at:

$$\mathbf{p} = \ell_1 \times \ell_2$$

$$\begin{aligned} &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & b_1 & c_1 \\ a_1 & b_1 & c_2 \end{vmatrix} = \mathbf{i}(b_1c_2 - b_1c_1) + \mathbf{j}(a_1c_2 - a_1c_1) + \mathbf{k} \cdot 0 \\ &= (b_1c_2 - b_1c_1, a_1c_2 - a_1c_1, 0) \end{aligned}$$

$(b_1c_2 - b_1c_1, a_1c_2 - a_1c_1, 0)$ is an ideal point. To summarise:

COROLLARY 1.6 (intersection of parallel lines in H.C.). Two parallel lines in H.C., $\ell_1 = (a, b, c)$ and $\ell_2 = (a, b, d)$, $c \neq d$ intersect at an ideal point \mathbf{p} (point at infinity) given by:

$$\mathbf{p} = \ell_1 \times \ell_2 = (b(d - c), a(d - c), 0) \quad (1.7)$$

In conclusion, very concept in projective geometry has its dual. If we start with a certain statement, e.g. “The line joining points $\mathbf{p}_1, \mathbf{p}_2$ is given by $\ell = \mathbf{p}_1 \times \mathbf{p}_2$ ” and take the dual of each concept, we end up with “The point of intersection of lines $\mathbf{l}_1, \mathbf{l}_2$ is given by $\mathbf{p} = \ell_1 \times \ell_2$ ”. For completeness, the duality table is given below.

Original	Dual
point	line
join	intersect
\mathbf{p}	ℓ

1.2.1 Geometric interpretation of conversions between E.C. and H.C.

So far 2D E.C. points were described as infinite rays in the 3D Euclidean space. In this section H.C. points are described in an alternative geometrical fashion – as straight 2D line segments. The most important thing to remember to interpret H.C. coordinates this way is that an infinite line through the origin is normal to a (3D) plane through the origin. In this section, the following are described in Euclidean space:

1. H.C. points,
2. a line intersection in H.C.,
3. and the points of the line joining two H.C. points

Below each case is derived geometrically.

1. **(H.C. point in E.C. space)** We already know that a H.C. point (ℓ_1, ℓ_2, ℓ_3) lies anywhere along the straight line from the origin to $\lambda(\ell_1, \ell_2, \ell_3)$, $\lambda \in \mathbb{R}$. The line defined by the direction of (ℓ_1, ℓ_2, ℓ_3) is normal to a plane through the origin that is normal to it, i.e. to plane $\ell_1 x_1 + \ell_2 x_2 + \ell_3 x_3 = 0$. Also, to find the Euclidean equivalent of a H.C. point, we set $x_3 = 1$. Ultimately, the Euclidean equivalent of the H.C. point (ℓ_1, ℓ_2, ℓ_3) is determined by the intersection of the following two planes:

$$\ell_1 x_1 + \ell_2 x_2 + \ell_3 x_3 = 0 \quad \cap \quad x_3 - 1 = 0$$

Therefore the 2D line equation that describes point (ℓ_1, ℓ_2, ℓ_3) is:

$$\ell_1 x_1 + \ell_2 x_2 + \ell_3 = 0$$

All this is visualised in Fig. 6 and Fig. 7.

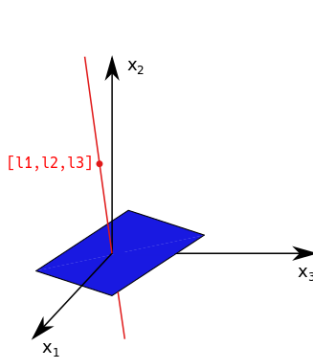


Fig. 6. The plane $\ell_1 x_1 + \ell_2 x_2 + \ell_3 x_3 = 0$

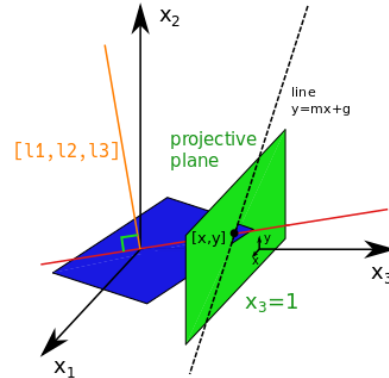


Fig. 7. The intersection of plane $\ell_1 x_1 + \ell_2 x_2 + \ell_3 x_3 = 0$ with $x_3 = 1$ describes H.C. point (ℓ_1, ℓ_2, ℓ_3) as a 2D straight line.

2. **(line intersection in H.C.)** We now know that a line along point $\ell = (\ell_1, \ell_2, \ell_3)$ in H.C. can be represented as the intersection of planes $\ell^\top \mathbf{x} = 0$, $x_3 = 1$ in the 3D space. Also the intersection of two E.C. points $\ell = (\ell_1, \ell_2, \ell_3)$, $\mathbf{m} = (m_1, m_2, m_3)$ can be described as an infinite ray spanned by the cross product $\ell \times \mathbf{m}$ (Cor. 1.3). Finally, to convert ray $\ell \times \mathbf{m}$ to its E.C. equivalent, we project it onto plane $x_3 = 1$

This is visualised in the figure below, where the red and blue planes are given, the orange ray describes their cross product, and the intersection of the orange ray with plane $x_3 = 1$ (which is a straight line) describes the intersection of the two lines defined by $\ell = (\ell_1, \ell_2, \ell_3)$ and $\mathbf{m} = (m_1, m_2, m_3)$ respectively. Note that the orange ray is NOT necessarily along the axis x_3 , although the figure is misleading.

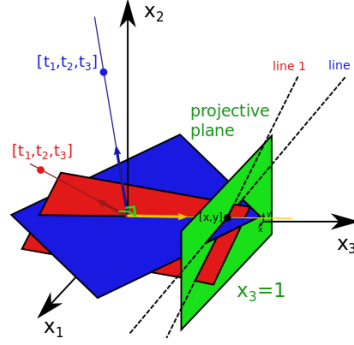


Fig. 8. Intersection of H.C. lines $\ell = (\ell_1, \ell_2, \ell_3)$ and $\mathbf{m} = (m_1, m_2, m_3)$ described as a 2D line.

3. **(Line joining two H.C. points)** It is known from Cor. 1.3 that if two H.C. points \mathbf{p}_1 , \mathbf{p}_2 are given, then the coefficients of the line through them in H.C. are given by the cross product $\ell = \mathbf{p}_1 \times \mathbf{p}_2$. This cross product is once again a vector normal to the plane $\ell_1 x_1 + \ell_2 x_2 + \ell_3 x_3 = 0$. To unproject the plane $\ell^\top \mathbf{x} = 0$ in 2D, we project it onto $x_3 = 1$. The line projected onto $x_3 = 1$ describes the line through \mathbf{p}_1 and \mathbf{p}_2 in 2D.

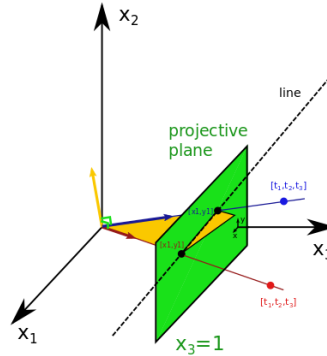


Fig. 9. The intersection of the plane $(\mathbf{p}_1 \mathbf{p}_2)^\top \mathbf{x} = 0$ with $x_3 = 1$ describes the line through H.C. points \mathbf{p}_1 and \mathbf{p}_2 .

1.3 The cross ratio

1.3.1 The maths of cross ratio

An important quantity that lets us convert scene (real world, 3D) distance measurements to image distance measurements (pixels) and vice versa is the cross ratio. The most important property of the cross ratio is that it remains invariant under the perspective projection.

DEFINITION 1.2 (cross ratio of points). The cross ratio of 4 collinear points A, B, C, D is defined as:

$$CR(A, B; C, D) = \frac{\frac{CA}{CB}}{\frac{DA}{DB}} \quad (1.8)$$

, where CA, CB, DA, DB are signed distances, i.e. each one depends on the direction of the line they lie at.

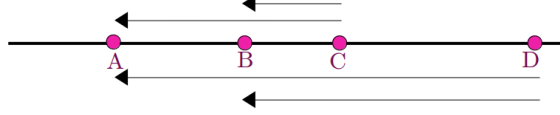


Fig. 10. The quantities that appear in the cross ratio $CR(A,B;C,D)$.

Obviously, the cross ratio does not depend on the direction of the line but it depends on the relative orientation of the 4 points. The cross ratio also does not depend on the choice of the projective coordinate system (it only involves distances). To remember Eq. (1.8), think of the double ratio fraction as a 4×2 matrix of 4 letters. On the first column ($CR(A,B;\dots)$), the letter changes every row. On the second column ($CR(\dots;C,D)$), it changes two rows.

The cross ratio has a number of interesting properties but they will not be discussed. We will only prove that it is invariant under the projective transform. However, before getting to the main point, it is helpful to the *cross ratio of 4 concurrent lines*.

DEFINITION 1.3 (cross ratio of lines). Given 4 concurrent lines a, b, c, d that meet a point M , the cross ratio of those lines as:

$$CR(a, b; c, d) = \frac{\frac{\sin(c\hat{M}a)}{\sin(c\hat{M}b)}}{\frac{\sin(d\hat{M}a)}{\sin(d\hat{M}b)}} \quad (1.9)$$

We can show that given 4 collinear points A, B, C, D and a centre of projection M , the cross ratio of points $CR(A, B; C, D)$ and the cross ratio of segments $a = MA, b = MB, c = MC, d = MD$ $CR(a, b; c, d)$ are equal.

LEMMA 1.2 (CR of points equals CR of segments). Given 4 collinear points A, B, C, D , then

$$CR(A, B; C, D) = CR(a, b; c, d) \iff \frac{\frac{CA}{CB}}{\frac{DA}{DB}} = \frac{\frac{\sin(c\hat{M}a)}{\sin(c\hat{M}b)}}{\frac{\sin(d\hat{M}a)}{\sin(d\hat{M}b)}} \quad (1.10)$$

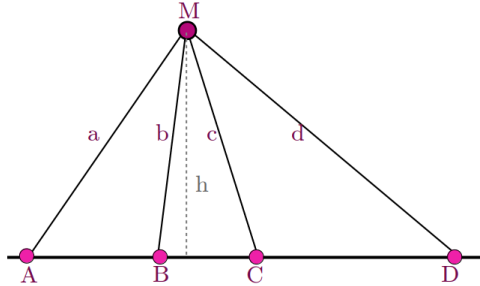


Fig. 11. 4 collinear points, their point of intersection M and the lines to M .

Proof. We express the area of each triangle in two different ways:

$$area(CMA) = \frac{1}{2} MC \cdot MA \sin(C\hat{M}A) = \frac{h}{2} CA$$

$$area(CMB) = \frac{1}{2} MC \cdot MB \sin(C\hat{M}B) = \frac{h}{2} CB$$

$$area(DMA) = \frac{1}{2} MD \cdot MA \sin(D\hat{M}A) = \frac{h}{2} DA$$

$$area(DMB) = \frac{1}{2} MD \cdot MB \sin(D\hat{M}B) = \frac{h}{2} DB$$

Then the cross ratio of distances is written as:

$$CR(a, b; c, d) = \frac{\frac{\sin(\hat{cMa})}{\sin(\hat{cMb})}}{\frac{\sin(\hat{dMa})}{\sin(\hat{dMb})}} = \frac{\frac{h/2 \cdot CA/(MC \cdot MA)}{h/2 \cdot CB/(MC \cdot MB)}}{\frac{h/2 \cdot DA/(MD \cdot MA)}{h/2 \cdot DB/(MD \cdot MB)}} = \frac{CA}{CB} \cdot \frac{DB}{DA}$$

□

We now have enough background to prove the main theorem.

THEOREM 1.1 (perspective transform preserves the cross ratio). *Any 4 points forming a pencil with common base have the same cross ratio. It is equal to the cross ratio of the base points.*

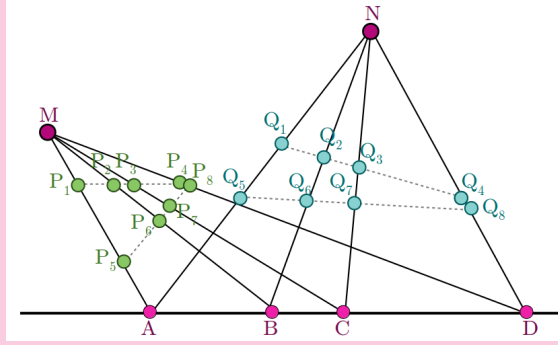


Fig. 12. $CR(P_1, P_2; P_3, P_4) = CR(P_5, P_6; P_7, P_8) = CR(Q_1, Q_2; Q_3, Q_4) = CR(Q_5, Q_6; Q_7, Q_8) = CR(A, B, C, D)$

For the sets of points in Fig. 12, the theorem states that:

$$CR(P_1, P_2; P_3, P_4) = CR(P_5, P_6; P_7, P_8) = CR(Q_1, Q_2; Q_3, Q_4) = CR(Q_5, Q_6; Q_7, Q_8) = CR(A, B, C, D) \quad (1.11)$$

In other words, the cross ratio remains invariant under projective transformations.

Proof. We proof the theorem for the green points first. We know from Eq. (1.10) that:

$$CR(P_1, P_2; P_3, P_4) = \frac{\frac{\sin(\hat{P}_3 \hat{M} P_1)}{\sin(\hat{P}_3 \hat{M} P_2)}}{\frac{\sin(\hat{P}_4 \hat{M} P_1)}{\sin(\hat{P}_4 \hat{M} P_2)}} = \frac{\frac{\sin(\hat{P}_7 \hat{M} P_5)}{\sin(\hat{P}_8 \hat{M} P_6)}}{\frac{\sin(\hat{P}_8 \hat{M} P_5)}{\sin(\hat{P}_8 \hat{M} P_6)}} = CR(P_5, P_6; P_7, P_8) \Rightarrow \quad (1.12)$$

$$CR(P_1, P_2; P_3, P_4) = CR(P_5, P_6; P_7, P_8) \quad (1)$$

, as P_i and P_{i+4} lie on the same line hence the sines are equal. In exactly the same way, it can be proven that:

$$CR(P_1, P_2; P_3, P_4) = CR(A, B, C, D) \quad (2)$$

Similarly:

$$CR(Q_1, Q_2; Q_3, Q_4) = CR(Q_5, Q_6; Q_7, Q_8) \quad (3)$$

From Eq. (1), (2), (3) we can derive the theorem:

$$CR(P_1, P_2; P_3, P_4) = CR(P_5, P_6; P_7, P_8) = CR(A, B, C, D) = CR(Q_1, Q_2; Q_3, Q_4) = CR(Q_5, Q_6; Q_7, Q_8)$$

□

1.3.2 Cross ratio and single view metrology

Cross ratio helps relate distances in the 3D world (xyz scene plane, e.g. in cm, m, or km) with distances in the image (xy image plane, e.g. in pixels). Inferring real-world measurements from image measurements is called single-view metrology.

Suppose we are given the image in pixels on the right and want to know how far the car is from the crossroad in the real world. We have measured that $A'D' = 300 \text{ px}$, $A'C' = 275 \text{ px}$ and $B'C' = 50 \text{ px}$. We want to know CD .

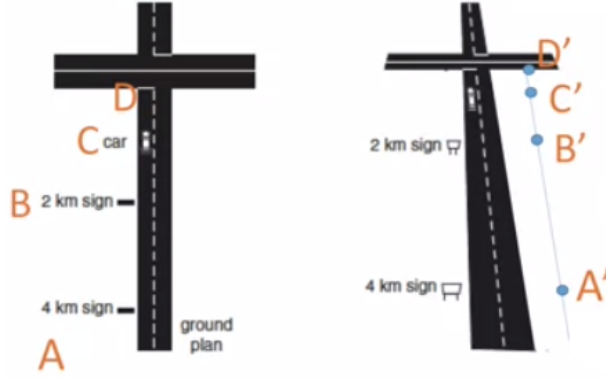


Fig. 13. An aerial view of a crossroad (left) and its image under the perspective transform of a camera (right).

A', B', C', D' are the projections of A, B, C, D on the ground plane therefore we can use the cross ratio preservation theorem:

$$\begin{aligned} \frac{A'C'}{A'D'} : \frac{B'C'}{B'D'} &= \frac{AC}{AD} : \frac{BC}{BD} \Rightarrow \\ \frac{275}{300} : \frac{50}{300 - 275} &= \frac{275}{300} : \frac{50}{75} = \frac{AC}{AD} : \frac{BC}{BD} \Rightarrow \\ \frac{AC}{AD} : \frac{BC}{BD} &= 1.375 \end{aligned}$$

However, $AC = 4 - CD$, $BC = 2 - CD$ therefore:

$$\frac{4 - CD}{4} : \frac{2 - CD}{2} = 1.375 \Rightarrow CD = 0.857 \text{ km}$$

If we know that a point in the scene is a vanishing point, i.e. at infinity, then inferring distances in the scene is even easier. For instance, in the picture below, point D' in the scene plane is at horizon of this plane, i.e. at infinity (also called vanishing point). When mapped to D in the image plane, it is at the horizon of the image, which is the vanishing line. The latter is a finite point. Note that D' in the image plane is at the intersection of lines $A'C'$ and ℓ , which are parallel in the real world but meet at a vanishing point in the image plane (Fig. 14).

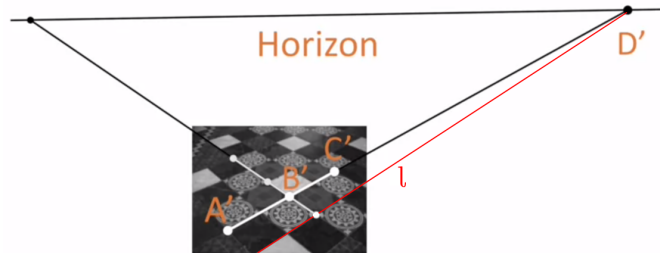


Fig. 14. Some square tiles in the image plane under perspective transform.

In the image plane, the vanishing point(s) can be determined as the intersection of two lines that are parallel in the real world.

Then, using the cross ratio theorem for the scene (points A, B, C, D , not shown) and image planes (Fig. 14) we get:

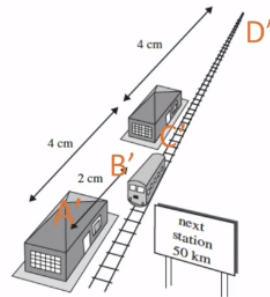
$$\frac{AC}{AD} : \frac{BC}{BD} = \frac{A'C'}{A'D'} : \frac{B'C'}{B'D'} =$$

However, $AD/BD = \|\mathbf{r}_A - \mathbf{r}_D\| / \|\mathbf{r}_B - \mathbf{r}_D\|$ and as $\mathbf{r}_D \rightarrow \infty$ then $\lim_{\mathbf{r}_D \rightarrow \infty} \|\mathbf{r}_A - \mathbf{r}_D\| / \|\mathbf{r}_B - \mathbf{r}_D\| = 1$ therefore the cross ratio in the scene plane is re-written as $\frac{AC}{AD} : \frac{BC}{BD} = \frac{AC}{BC}$.

COROLLARY 1.7 (cross ratio including vanishing point). *If A, B, C, D are some scene (real-world) points and D is a vanishing point, i.e. at infinity, then their cross ratio can be simplified such that it does not take D into account:*

$$\frac{A'C'}{A'D'} : \frac{B'C'}{B'D'} = \frac{AC}{BC}, \quad \mathbf{r}_D \rightarrow \infty \quad (1.13)$$

Revisiting the train position inference problem, knowing that a point is vanishing allows to measure distances in the real world with only one reference distances.



How far away is the train from the next station?
Or $BC = ?$

$$\frac{A'C'}{A'D'} : \frac{B'C'}{B'D'} = \frac{AC}{BC}$$

$$\frac{4}{8} : \frac{2}{6} = \frac{50}{BC}$$

$$\frac{4}{8} : \frac{2}{6} = \frac{50}{BC} \text{ and } BC = 33.33 \text{ km.}$$

Fig. 15. To find the position of the train in the real world, we only need a vanishing point and the distance $A'D'$.

1.3.3 Cross ratio and height forensics

Cross ratio given a vanishing point, as in Eq. (1.13), is the main step of inferring the height of an object in an image given the height of another object as reference. Before explain how this is done, remember that points in parallel lines in the scene plane lie on the *same* line in the image plane.

Suppose the we want to infer the height of the man in black hat in the image below given the height of the statue.

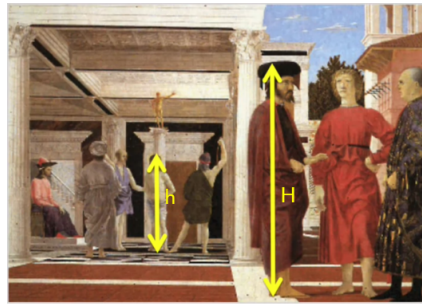


Fig. 16. The height H of a person to be determined given the reference height h of the statue.

We break the solution in several steps. Dashed points denote the scene and non-dashed the image.

1. **(Find the vanishing line in the image).** To find the vanishing line, take the intersection of two lines that are obviously parallel in the scene. The horizontal line through that intersection is the vanishing line (horizon). All vanishing points lie along that line.

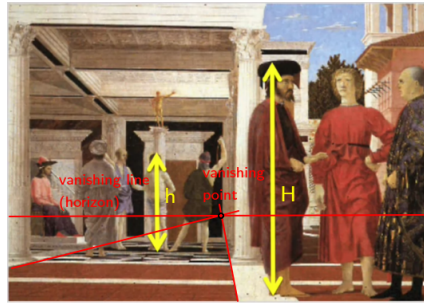


Fig. 17. Finding the vanishing line.

2. **(Find a vanishing point)** We then find a vanishing point defined by the objects of interest. For that, we need a line in the real world that connects two parallel objects – in this case the feet. We intersect that line with the horizon to find the vanishing point A.

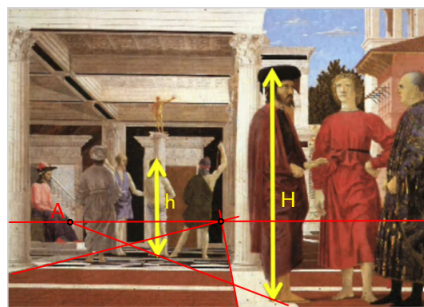


Fig. 18. Finding the vanishing line.

3. **(Distance transfer)** We know $h = B'C'$ in the scene plane, e.g. 179 cm. We transfer the known height h onto the unknown object. Draw one more line from the vanishing point A through the top C of the known object and intersect it with the vertical of the unknown object at E . Because A, C, E or on the same line in the image, they're parallel in the scene. Likewise for A, B, D . Therefore in the scene $D'E' = B'C' = h$. But $D'E'$ is not the full height of the unknown object. We know $D'E'$ and want $D'F'$.

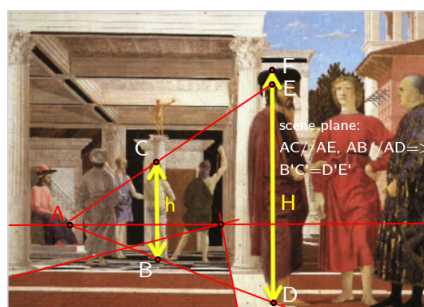


Fig. 19. Transferring the height of the known object onto the unknown one.

4. **(Find the vertical vanishing point and use the cross ratio)** To find the height of the man, we want to know $D'E'/D'F'$. If the horizon is not in the middle, lines BC and DF are not actually parallel in the image plane, but they meet at a finite point V . V is therefore the intersection of BC and DF .

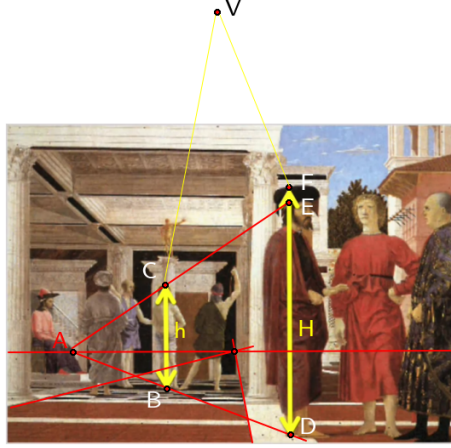


Fig. 20. Lines BC and DF intersect at the vertical vanishing point V .

V' is at infinity in the scene plane. Therefore if we use Eq. (1.13) then we can determine ratio $D'E'/D'F'$ (knowing $D'E' = h$):

$$\frac{D'E'}{D'F'} = \frac{DE}{DF} : \frac{VE}{VF} \Rightarrow D'F' = \frac{1}{D'E'} \frac{VE}{VF} : \frac{DE}{DF} \Rightarrow H = \frac{1}{h} \frac{VE}{VF} : \frac{DE}{DF}$$

COROLLARY 1.8 (height measurement via distance transfer). *If we are given an image of a reference object of height $h = B'C'$ in the scene, then we can determine the height H (in the scene) of an unknown object by the formula:*

$$H = \frac{1}{h} \frac{VE}{VF} : \frac{DE}{DF} \quad (1.14)$$

Points V, E, F, D are marked in Fig. 20.

These are the core steps height measurement via perspective. Additionally, point V can be cross-produced as $\mathbf{v} = \overrightarrow{DF} \times \overrightarrow{BC}$ as in Cor. 1.2, finding the horizon can be automated by Hough lines, or the algorithm can be semi-supervised by allowing the user to click at pixels in order to define lines.

A Appendices

A.1 Pinhole camera model derivation

Consider the case of a *pinhole camera*, which consists of a box which a tiny aperture on the front plane and a film which captures the energy (colour) of the light on the real plane (Fig. 21). We assume that all light rays entering the camera meet at the aperture (centre of projection). The horizontal distance between the aperture and the film is called focal length f .

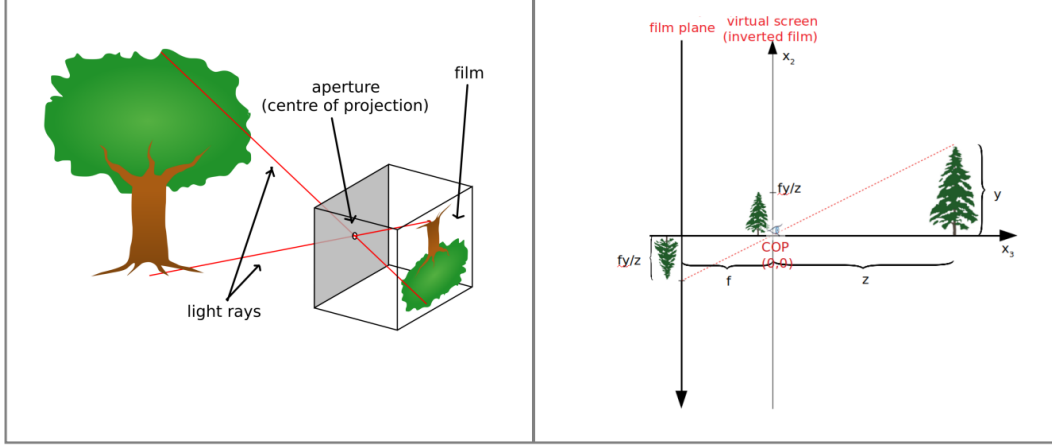


Fig. 21. Left: Pinhole camera schematic in 3D. Right: Pinhole camera model of a tree in the y dimension.

From the similar triangles in front and behind the screen in Fig. 21, we obtain for the project y' of the object:

$$\frac{y}{z} = \frac{y'}{f} \Rightarrow y' = f \frac{y}{z}$$

Similarly for the projected x' (not shown in Fig. 21) we obtain:

$$x' = f \frac{x}{z}$$

To reiterate, pinhole camera performs the following linear mapping in Cartesian coordinates:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} \mapsto \begin{bmatrix} \frac{fx}{z} \\ \frac{fy}{z} \\ \frac{f}{z} \end{bmatrix}, \quad \begin{bmatrix} \frac{fx}{z} \\ \frac{fy}{z} \\ \frac{f}{z} \end{bmatrix} = \frac{1}{z} \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

The position (x', y') on the screen is measured in distance units – not pixels.