CIS 5800

Machine Perception

Or *Geometric Computer Vision*

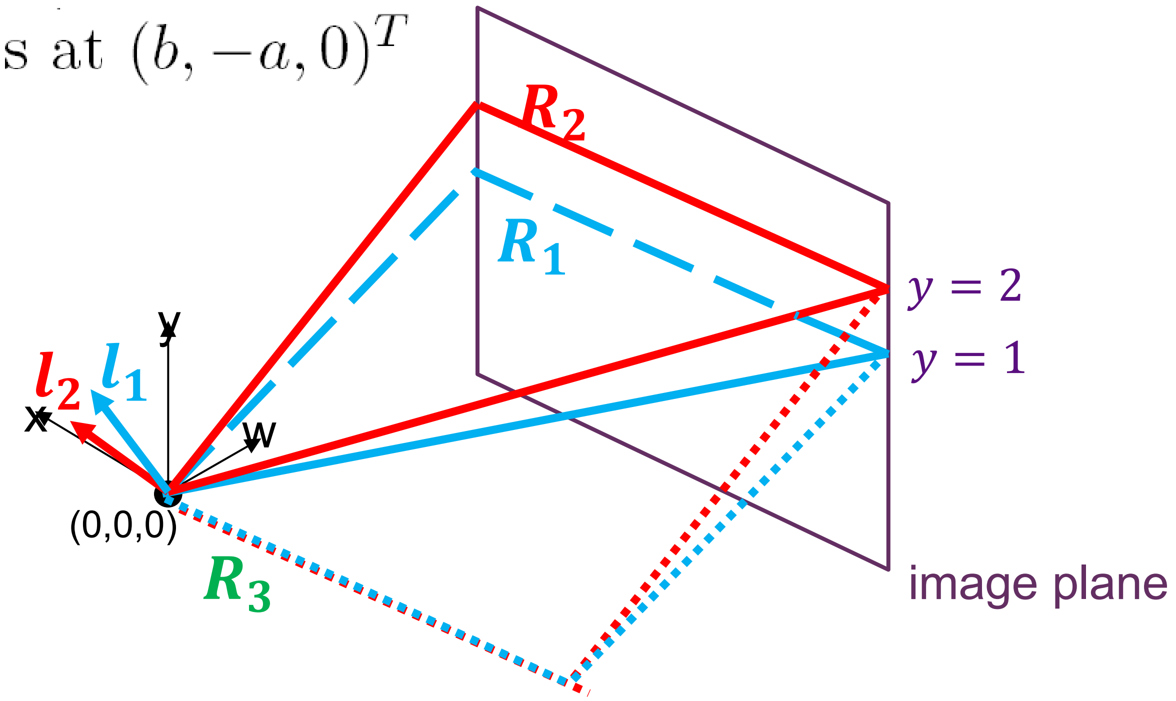
Instructor: Lingjie Liu

Lec 4: Jan 29, 2025

Recap: Extending to Any Parallel Lines

- Under projective geometry,
  - All parallel lines intersect at a point at infinity
  - One point at infinity ⇔ one parallel line direction

$$\text{line } l = (a, b, c)^T \text{ intersects at } (b, -a, 0)^T$$

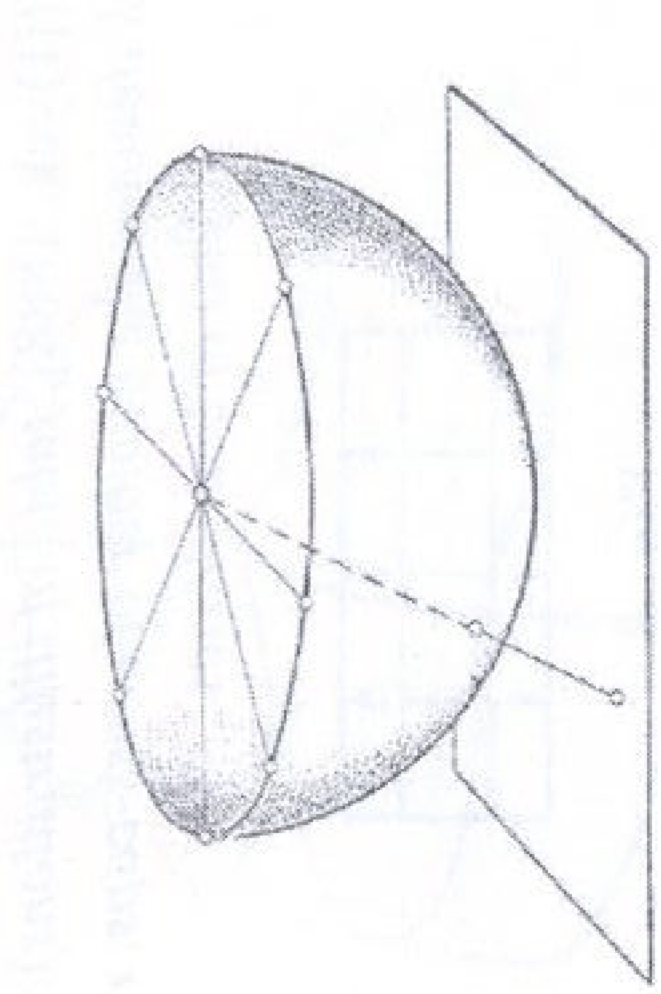


Recap: Point at infinity / “ideal” points

$$(\underbrace{x_1,x_2}, 0)$$

Looking-at direction

“Ideal” points



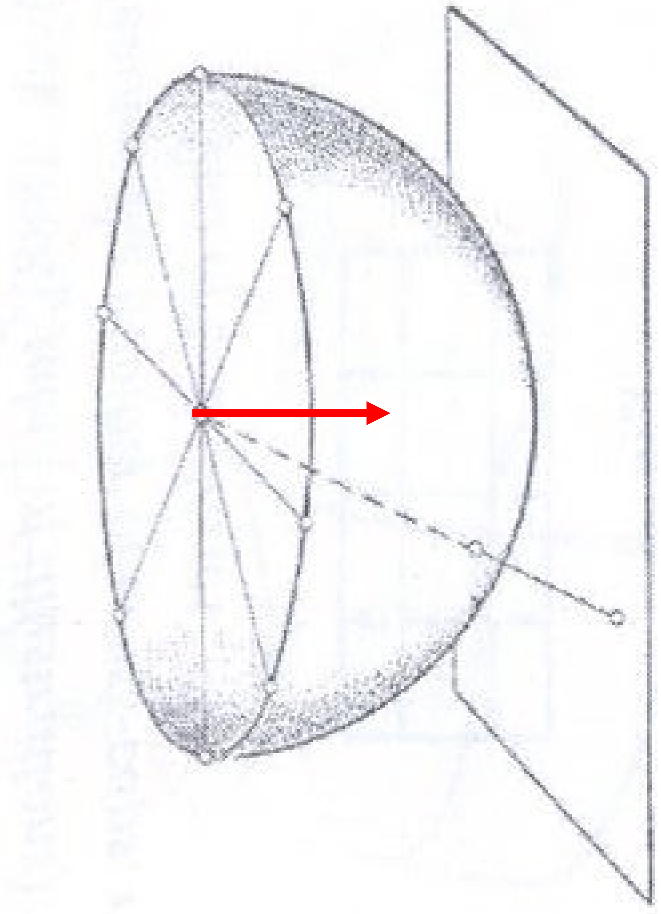
Recap: “Line at infinity”

- A line passing through all ideal points i.e. point

$$l_\infty = (0,0,1)^T$$

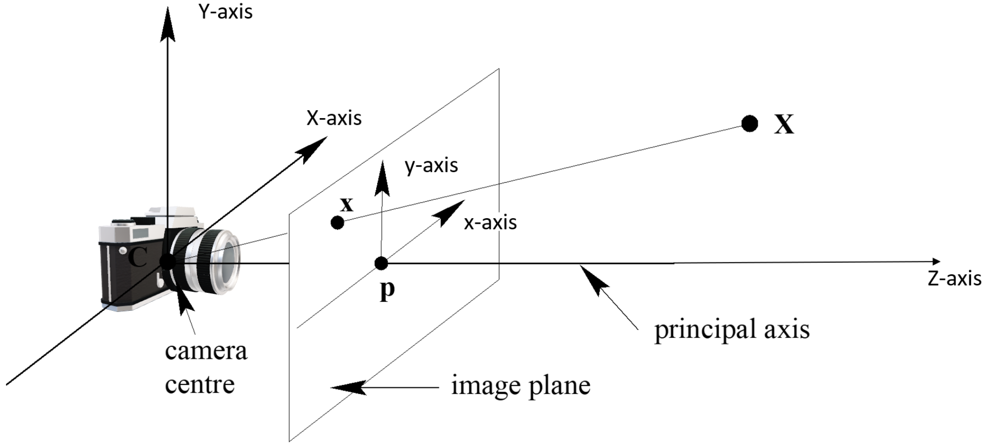
  - Because :

$$\begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ 0 \end{bmatrix} = 0$$



Recall: Camera Projection Equation

$$x = f\frac{X}{Z}, y = f\frac{Y}{Z}$$

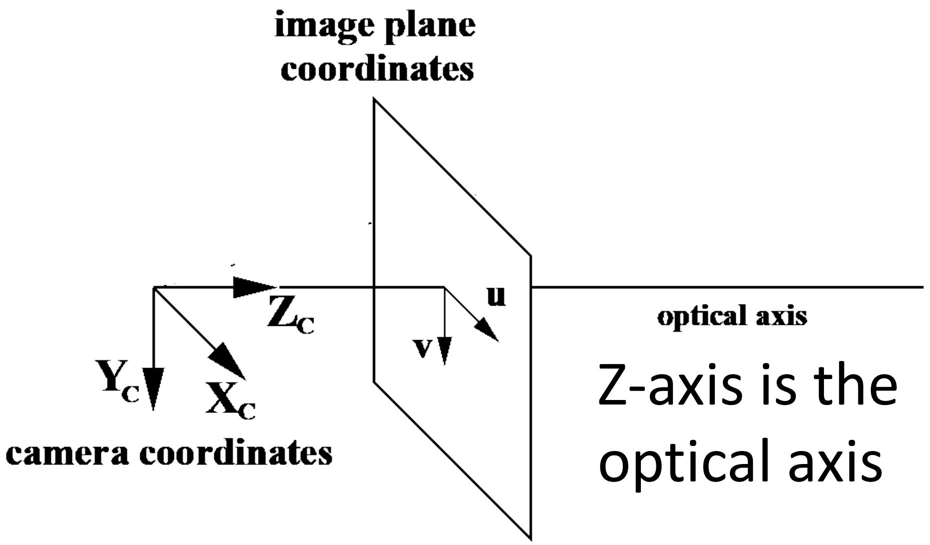


$$\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \rightarrow \begin{bmatrix} fX \\ fY \\ Z \end{bmatrix} = \begin{bmatrix} f & \square & \square & 0 \\ \square & f & \square & 0 \\ \square & \square & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

This assumes that:

- the image coordinate system origin is the same as the "principal point" p where the principal / optical axis intersects image plane. Sometimes called "image center"
- Points in the 3D world are known in the *camera-centric* coordinate system.

Camera Coordinate System + Principal Point Offset



Z-axis is the optical axis

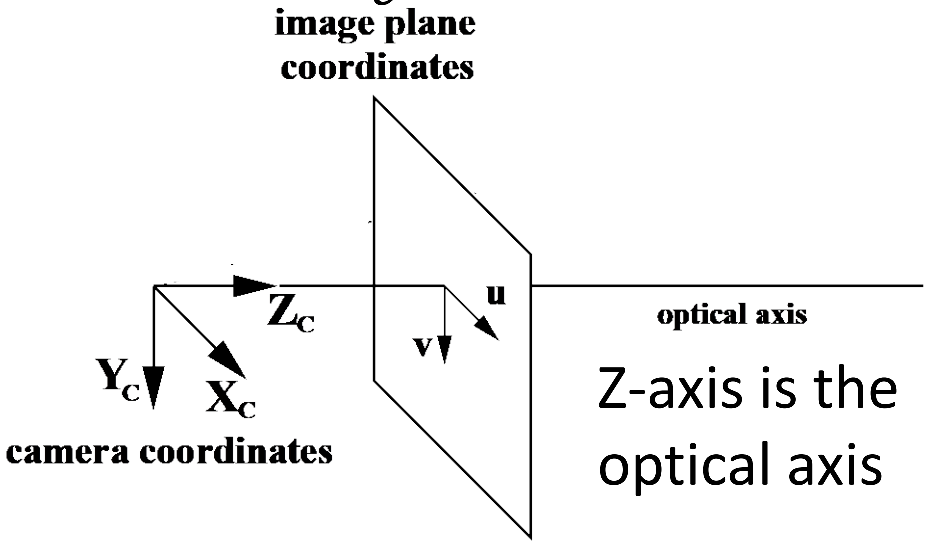
The image plane (u, v) is perpendicular to the optical axis.
Intersection of the image plane with the optical axis is the *image center* (u_o, v_o)

Projection in pixels

$$u = f\frac{X_c}{Z_c} + u_0, \qquad v = f\frac{Y_c}{Z_c} + v_0$$

Projection equation with image origin ≠ principal point

$$u = f\frac{X_c}{Z_c} + u_0, \qquad v = f\frac{Y_c}{Z_c} + v_0$$

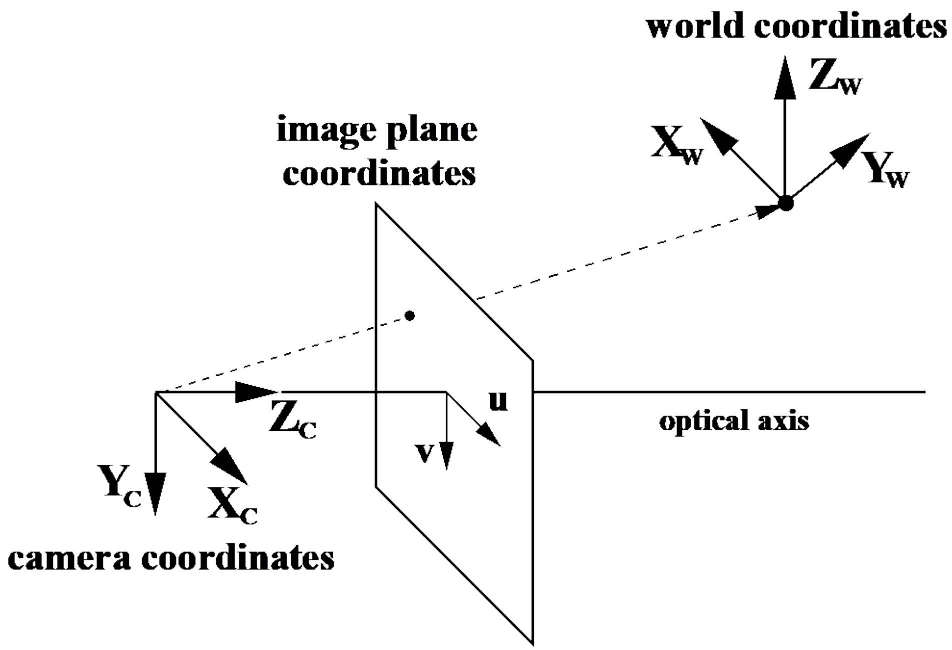


Z-axis is the optical axis

Bold face to indicate this is the full vector in $\mathbb{P}^3$

$$\lambda\begin{bmatrix}u\\v\\1\end{bmatrix} = \begin{bmatrix}f & & u_0\\ & f & v_0\\ & & 1\end{bmatrix}\begin{bmatrix}1 & & & 0\\ & 1 & & 0\\ & & 1 & 0\end{bmatrix}\begin{bmatrix}X_c\\Y_c\\Z_c\\1\end{bmatrix} = K[I|0]\boldsymbol{X}_c$$

From world to camera: Euclidean transformation



$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = R_{3\times 3} \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} + t = \begin{bmatrix} R_{3\times 3} & \boldsymbol{t} \\ 0 & 1 \end{bmatrix} \boldsymbol{X}_w$$

What do R and t mean exactly?

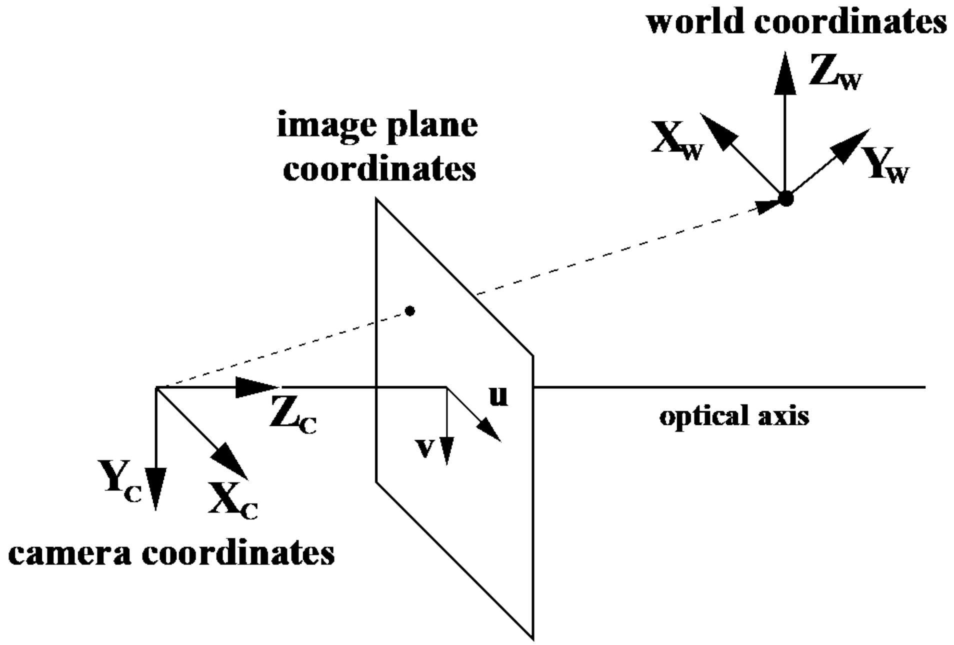
R denotes the rotation of the world axes w.r.t. camera axes

= *inverse* rotation of the camera axes w.r.t. world axes.

What about $\boldsymbol{t}$?

If $\boldsymbol{a}$ were the translation of the world origin from the camera origin, then $R(x + a)$ would be the camera coords of a world point x. i.e. $Rx + Ra$. So $\boldsymbol{t}$ here is actually R times the translation of world origin from camera origin.

Putting the pieces together: Projection matrix P



$$\lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \underbrace{\begin{bmatrix} f & & u_0 \\ & f & v_0 \\ & & 1 \end{bmatrix}}_{\text{camera 3D coords to pixels}} \underbrace{[R_{3\times 3} \quad \boldsymbol{t}] \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}}_{\text{Convert world to camera coordinates}} = K[R|\boldsymbol{t}]\boldsymbol{X}_w$$

Summarizing

- Permitting principal point offset (i.e. shifted origin for image plane)

$$\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \rightarrow \begin{bmatrix} fX + Zu_0 \\ fY + Zv_0 \\ Z \end{bmatrix} = \underbrace{\begin{bmatrix} f & & u_0 \\ & f & v_0 \\ & & 1 \end{bmatrix}}_{\text{“Intrinsics” } K} \begin{matrix} 0 \\ 0 \\ 0 \end{matrix} \Bigg] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = K[I|\mathbf{0}]\boldsymbol{X_c}$$

- The above assumes X, Y, Z in camera-centric coordinates. Permitting different “world” and “camera” coordinate frames:

$$X_c = \underbrace{R_{3\times3}}_{\text{Rotate}} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} + \underbrace{\boldsymbol{t}}_{\text{Translate}} = [R|\boldsymbol{t}]\boldsymbol{X_w},$$

$$\boxed{\boldsymbol{x} \sim K[R|\boldsymbol{t}]\boldsymbol{X_w}}$$

First convert to camera coordinates

Then project as before.

Perspective Projection in homogeneous coordinates

$$\begin{bmatrix} x \\ y \\ w \end{bmatrix} = P_{3\times4} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

$$\underbrace{P_{3\times4}}_{\text{“Camera projection matrix”}} = \underbrace{K_{3\times3}}_{\text{“Intrinsics”}} \underbrace{[R_{3\times3} | \boldsymbol{t}_{3\times1}]}_{\text{“Extrinsics”}}$$

Q: How many degrees of freedom does P have given the type of intrinsics we have assumed?

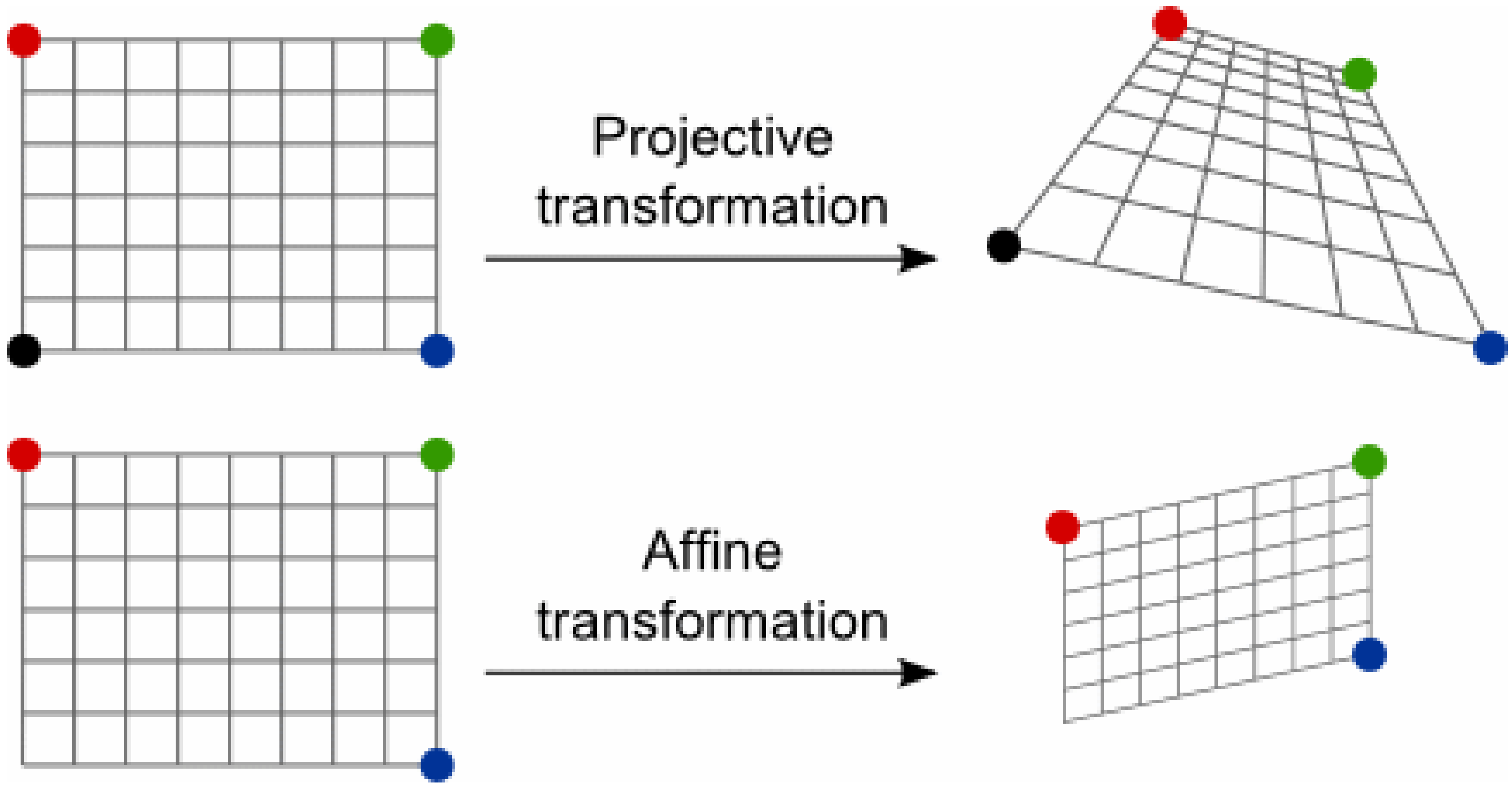
Projective Transformations

aka Collineations
aka Homographies
aka Projectivity

Example of Projective Transformation

Common notations: H

(Note that some books use A; however, we will avoid using A in this course, as A is commonly associated with Affine Transformations.)



Example of Projective Transformation

- A 2D point before H is represented as (X, Y) , after Projective transformation is (u, v) :

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \sim H \begin{pmatrix} X \\ Y \\ 1 \end{pmatrix}$$

or

$$\begin{aligned} u &= \frac{H_{11}X + H_{12}Y + H_{13}}{H_{31}X + H_{32}Y + H_{33}} \\ v &= \frac{H_{21}X + H_{22}Y + H_{23}}{H_{31}X + H_{32}Y + H_{33}} \end{aligned}$$

Projective Transformation = Homography = Collineation=Projectivity

Definition

A **projective transformation** is any invertible matrix transformation $\mathbb{P}^2 \rightarrow \mathbb{P}^2$.

Projective Transformation = Homography = Collineation=Projectivity

Definition

A **projective transformation** is any invertible matrix transformation $\mathbb{P}^2 \to \mathbb{P}^2$.

A projective transformation H maps p to $p' \sim Hp$

Invertibility means that $\det(H) \neq 0$ and that there exists $\lambda \neq 0$ such that $\lambda p' = Hp$

Observe that we will write either $p' \sim Hp$ or $\lambda p' = Hp$

How many unknowns are in a projective transformation H ?

$(\mathbb{P}^2 \rightarrow \mathbb{P}^2)$

A projective transformation $\mu\ H$ is the same as H since they map to projectively equivalent points:

$$\mu\lambda p' = \mu\ Hp$$

How many unknowns are in a projective transformation H ?

$(\mathbb{P}^2 \rightarrow \mathbb{P}^2)$

A projective transformation $\mu\, H$ is the same as H since they map to projectively equivalent points:

$$\mu \lambda p' = \mu \; Hp$$

We will be able to determine a projective transformation only up to a scale factor. Hence the 3x3 invertible matrix H will have only EIGHT independent unknowns.

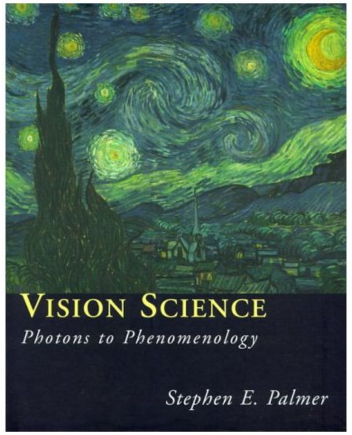
Q: How many unknowns in a projective transformation in $\mathbb{P}^3$?

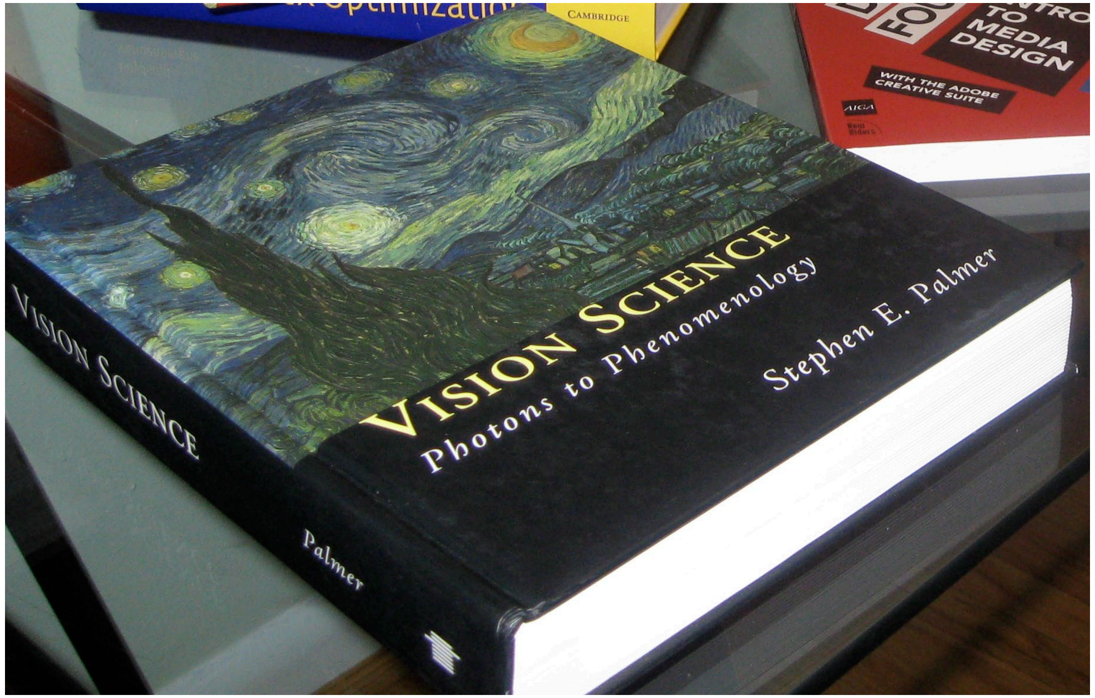
Perspective Projection v.s. Projective Transformation

	Perspective Projection	**Projective Transformation**
Definition	A mapping from 3D space to a 2D plane (e.g., camera image)	A general mapping between projective space (e.g., P^2 to P^2)
Mathematical Formula	p′=K[R|T]P	p′=H·p
Input Space	R^3 (can also be P^3)	P^n (typically P^2 in this class)
Output Space	R^2 (can also be P^2)	P^n (typically P^2 in this class)
Applications	Image formulation, 3D rendering	Image registration, planar transformation, texture mapping

When Perspective Projection -> Projective Transformation?

A perspective camera projection of a plane (i.e., a camera image) is a projective transformation in $\mathbb{P}^2$





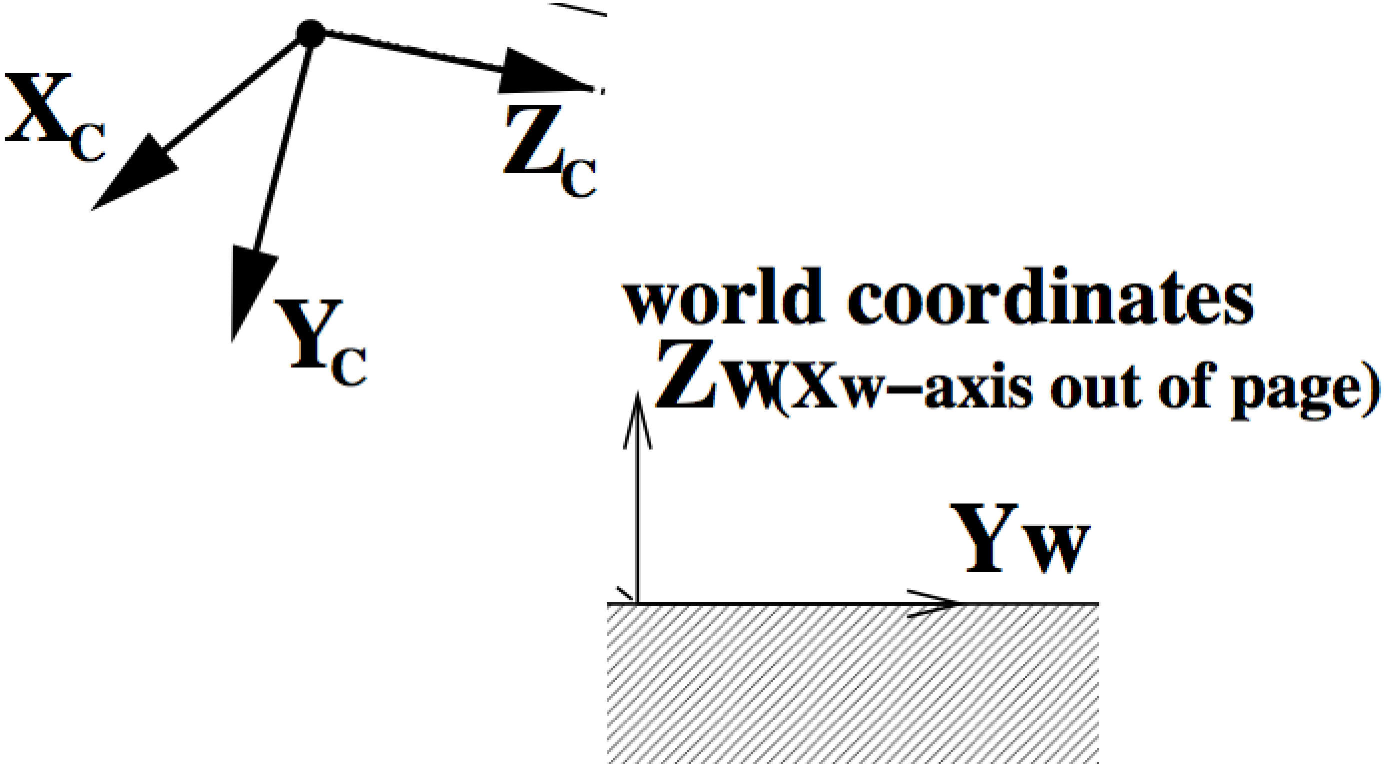
When Perspective Projection -> Projective Transformation?

- Can we show that the perspective camera projection from $\mathbb{P}^3 \to \mathbb{P}^2$ of a plane in the world is in fact a homography in $\mathbb{P}^2$ (i.e., projective transformation from $\mathbb{P}^2 \to \mathbb{P}^2$) when the world plane coordinates are expressed in $\mathbb{P}^2$?

- Remember:

$$\begin{bmatrix} x \\ y \\ w \end{bmatrix} \sim K_{3\times3}[R_{3\times3}|\boldsymbol{t}_{3\times1}]\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

Assume world plane $Z_w = 0$



When Perspective Projection -> Projective Transformation?

Recall the projection from world to camera

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} \sim K \begin{pmatrix} r_1 & r_2 & r_3 & T \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ W \end{pmatrix}$$

and assume that all points in the world lie in the ground plane $Z = 0$.

Pose From Homography

Recall the projection from world to camera

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} \sim K \begin{pmatrix} r_1 & r_2 & r_3 & T \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ W \end{pmatrix}$$

and assume that all points in the world lie in the ground plane $Z = 0$.

Then the transformation reads

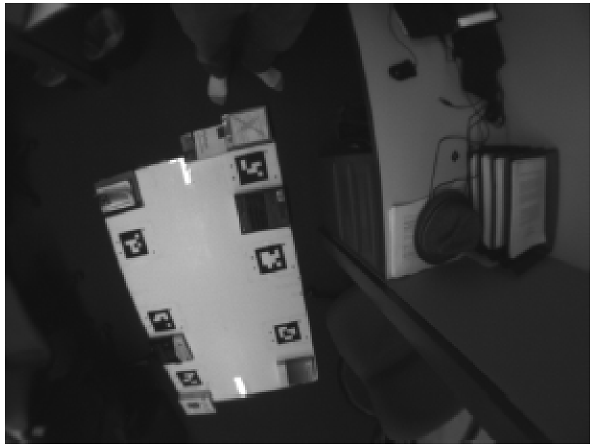
$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} \sim \underbrace{K \begin{pmatrix} r_1 & r_2 & T \end{pmatrix}}_{\substack{\text{The planar homography} \\ H: \mathbb{P}^2 \rightarrow \mathbb{P}^2}} \begin{pmatrix} X \\ Y \\ W \end{pmatrix}$$

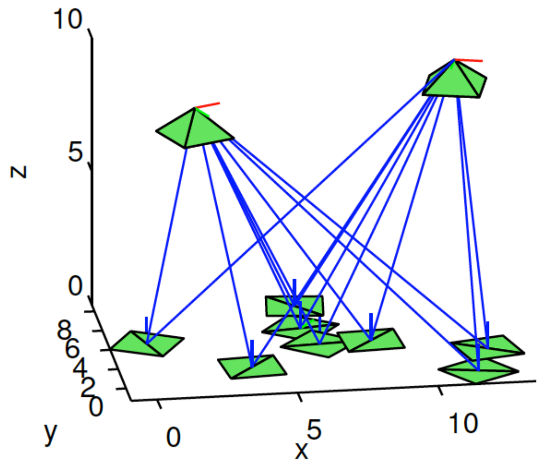
Computing the homography can tell us how the camera (and therefore, e.g. a robot attached to the camera) is oriented w.r.t. to a world plane! (assuming known K)

Q: Where do you get $\boldsymbol{r}_3$ from though?

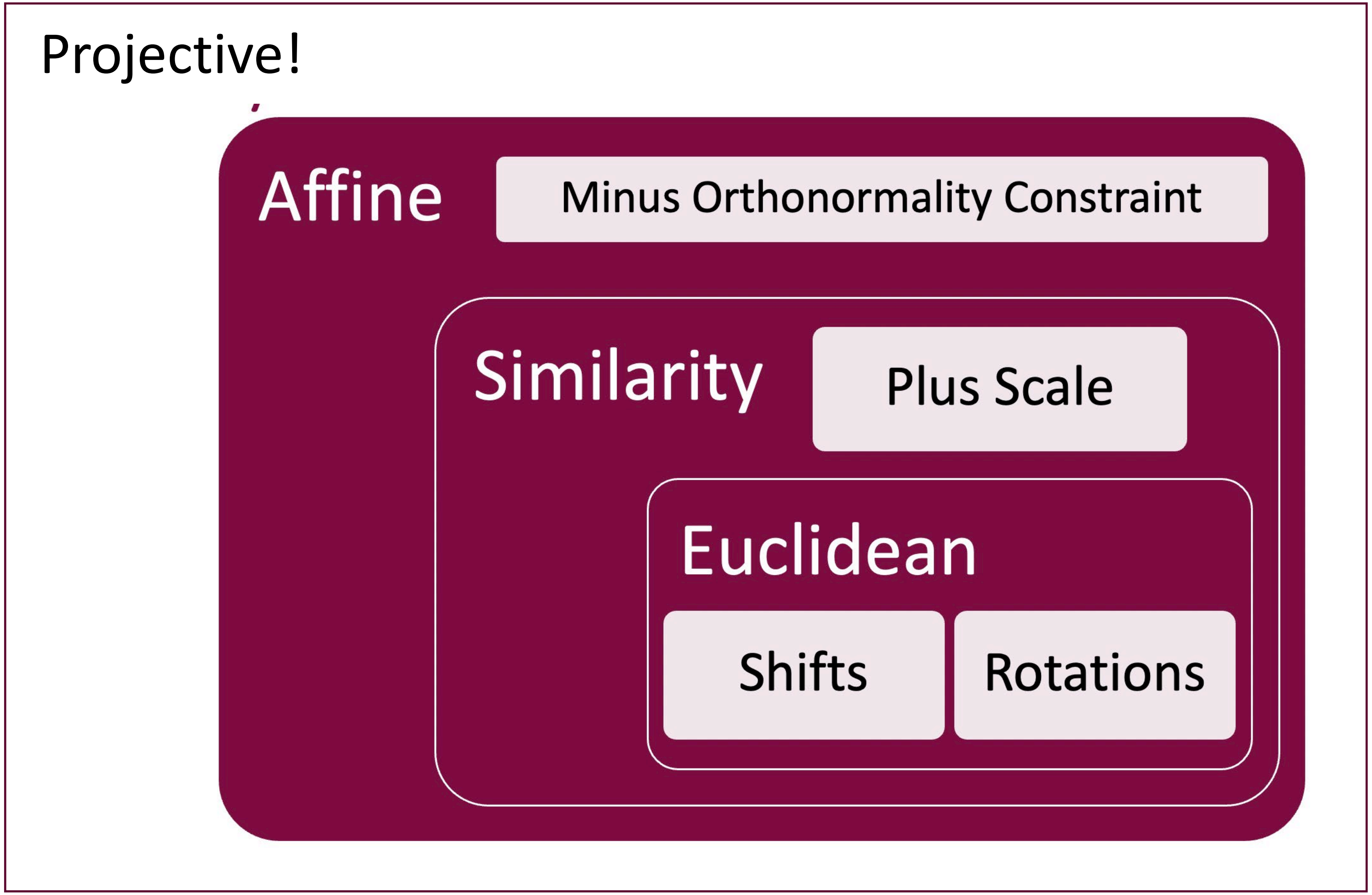
A: $\boldsymbol{r}_3 = \boldsymbol{r}_1 \times \boldsymbol{r}_2$

Localization w.r.t. known planes using homographies



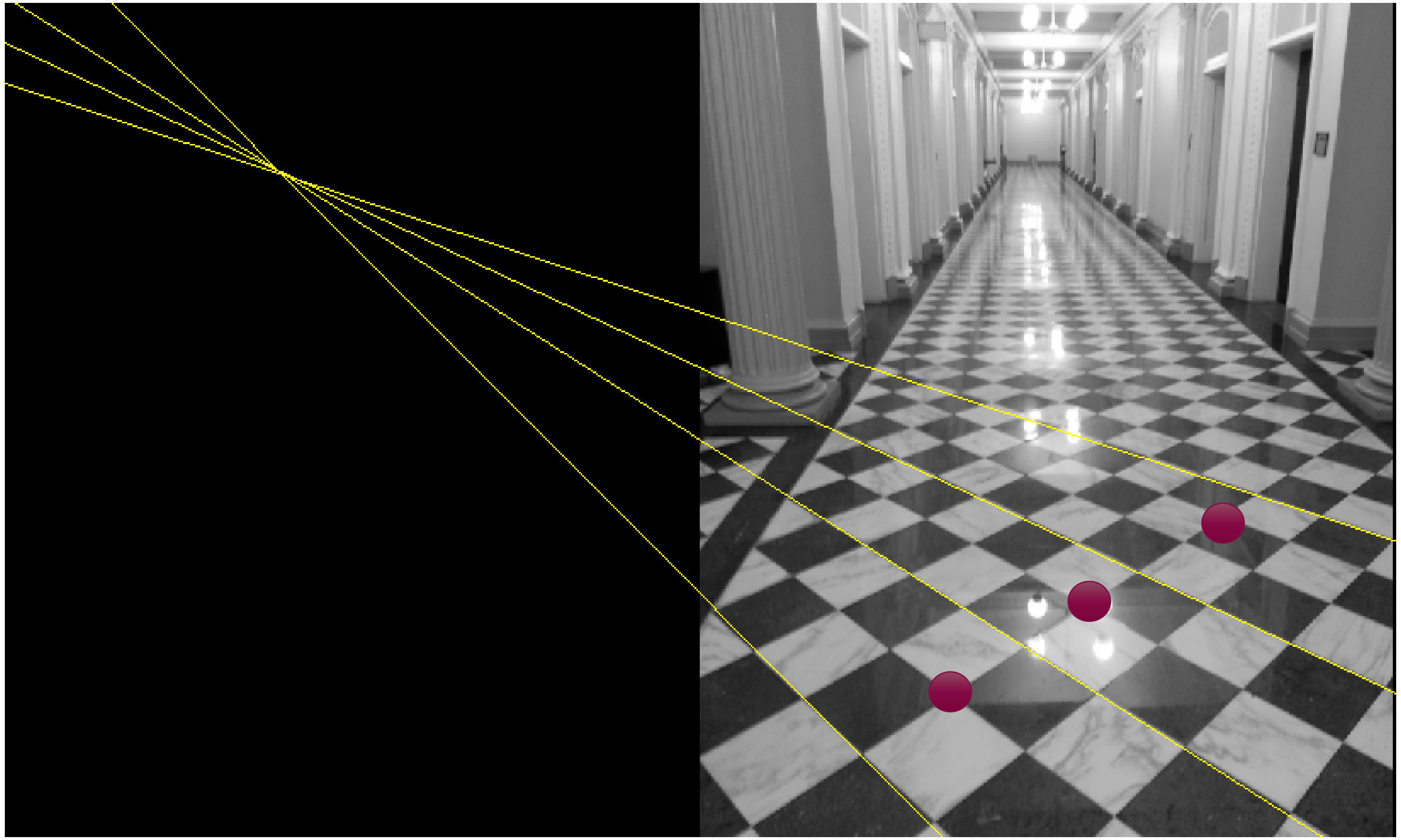


Place in the Hierarchy of Transformations

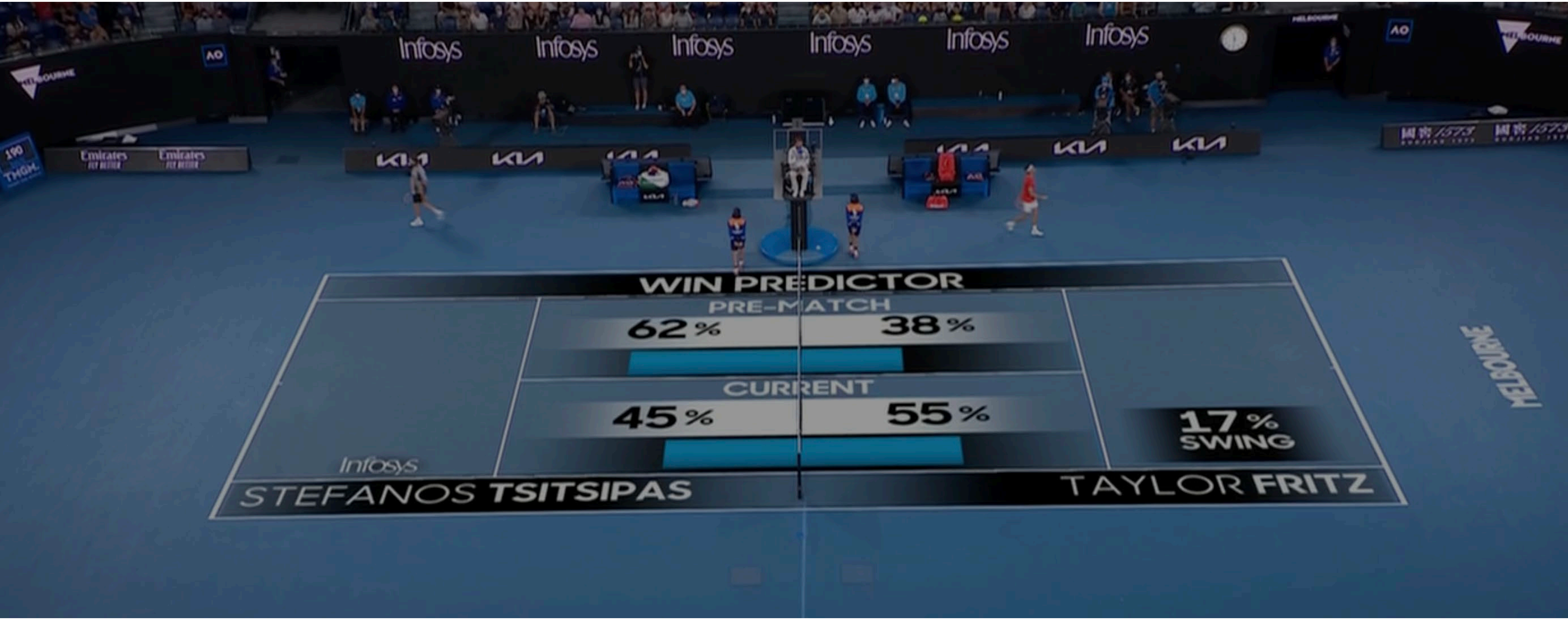


A projective transformation preserves incidence:

- Three collinear points are mapped to three collinear points.
- and three concurrent lines are mapped to three concurrent lines.



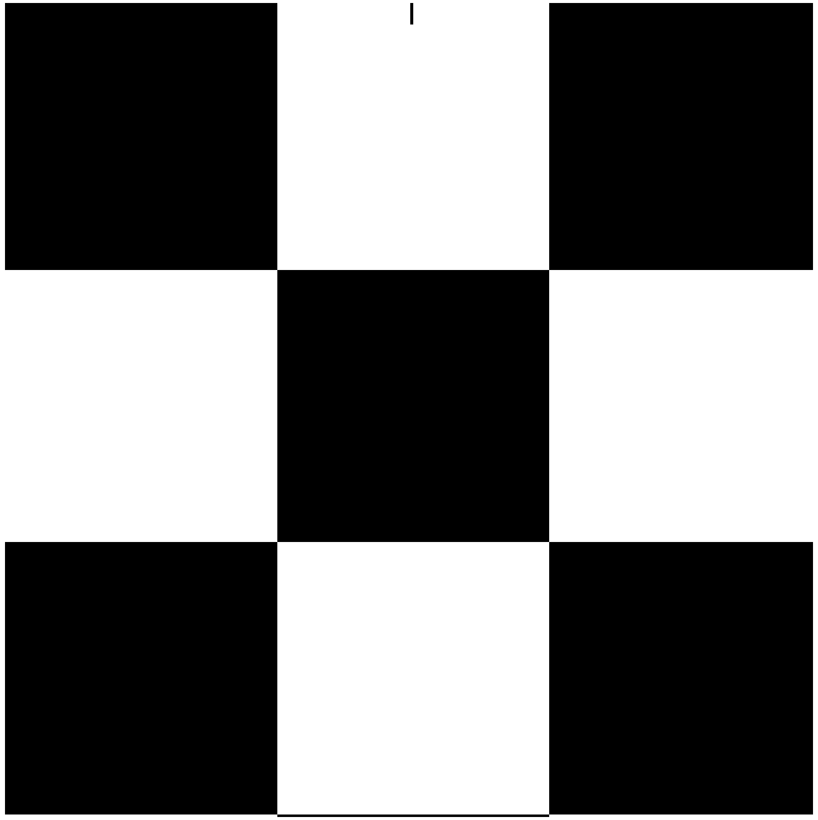
Application Example: Virtual Billboards



Computing Homographies From 4 Point Correspondences

“4-point collineation”

How can we compute the projective transformation between a known pattern and its projection?



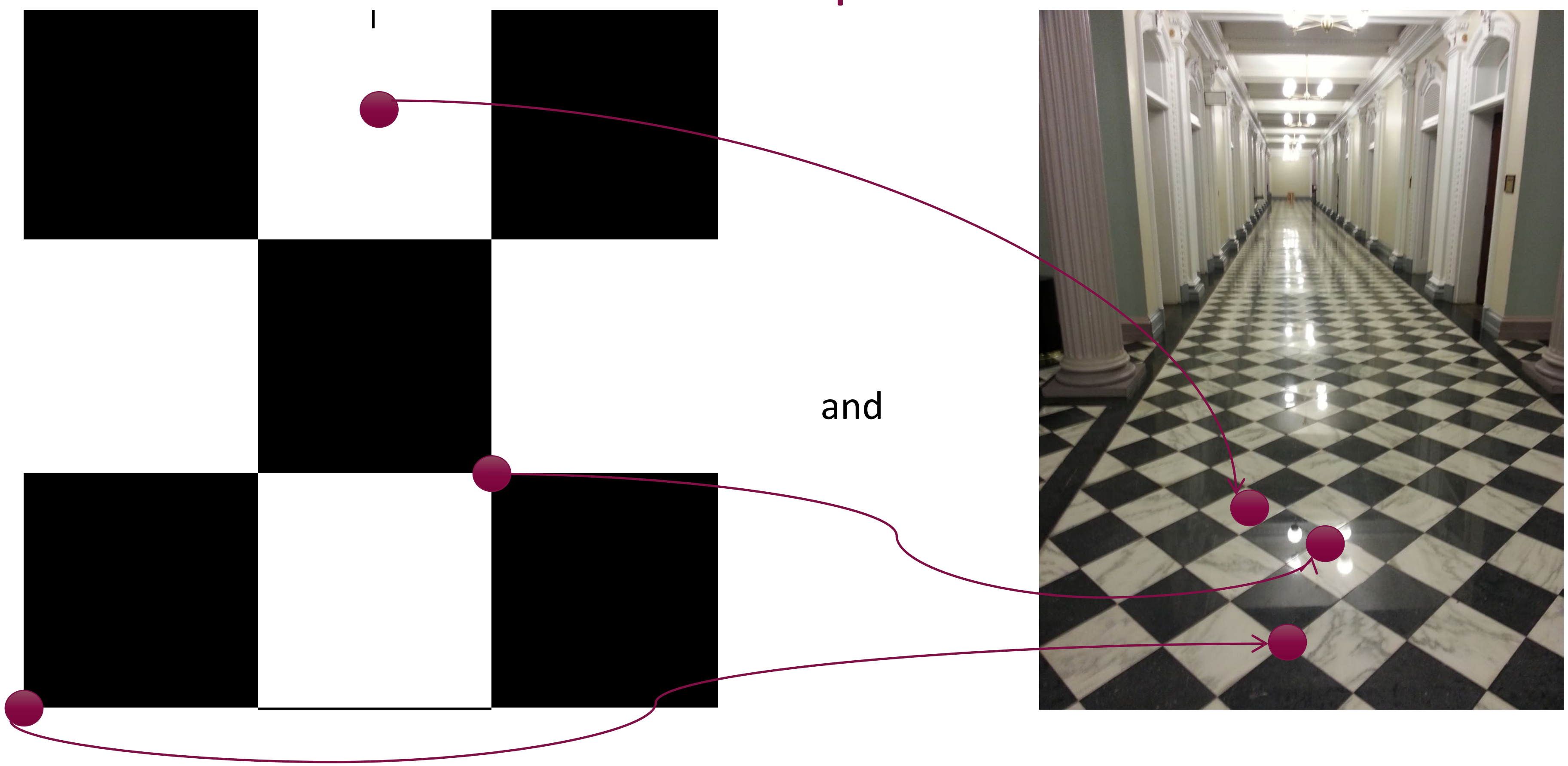
Floor tiles measured in [m]

and

Points in pixel coordinates

The result of such a transformation would map any point in one plane to the corresponding point in the other

“correspondences”



Floor tiles measured in [m]

Points in pixel coordinates

Recap:

How many unknowns are in a projective transformation H ?

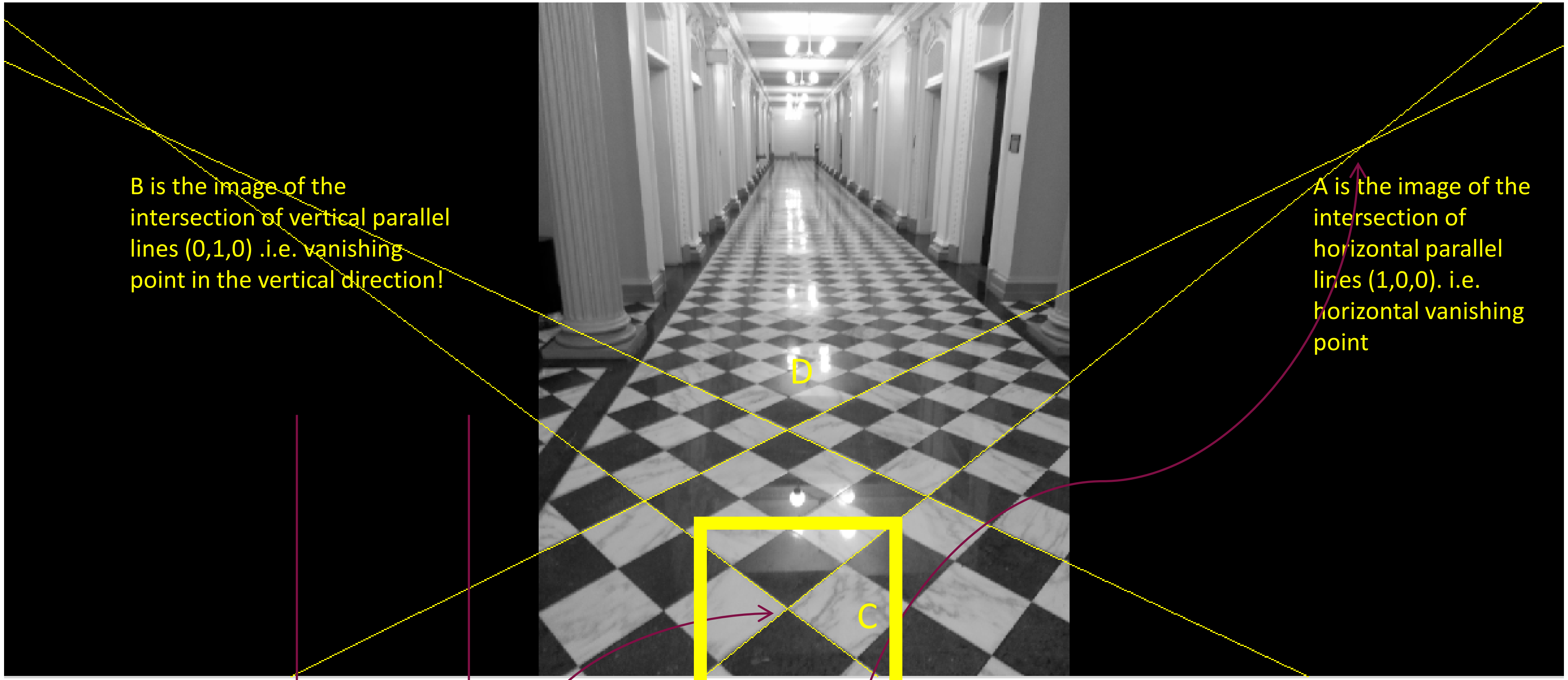
$(\mathbb{P}^2 \rightarrow \mathbb{P}^2)$

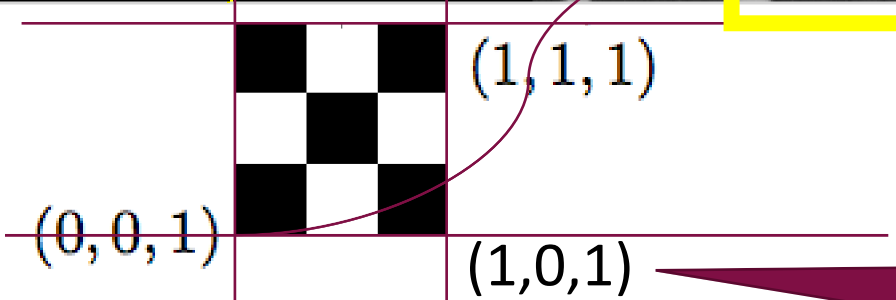
A projective transformation μH is the same as H since they map to projectively equivalent points:

$$\mu\lambda p' = \mu\ Hp$$

We will be able to determine a projective transformation only up to a scale factor. Hence the 3x3 invertible matrix H will have only EIGHT independent unknowns.

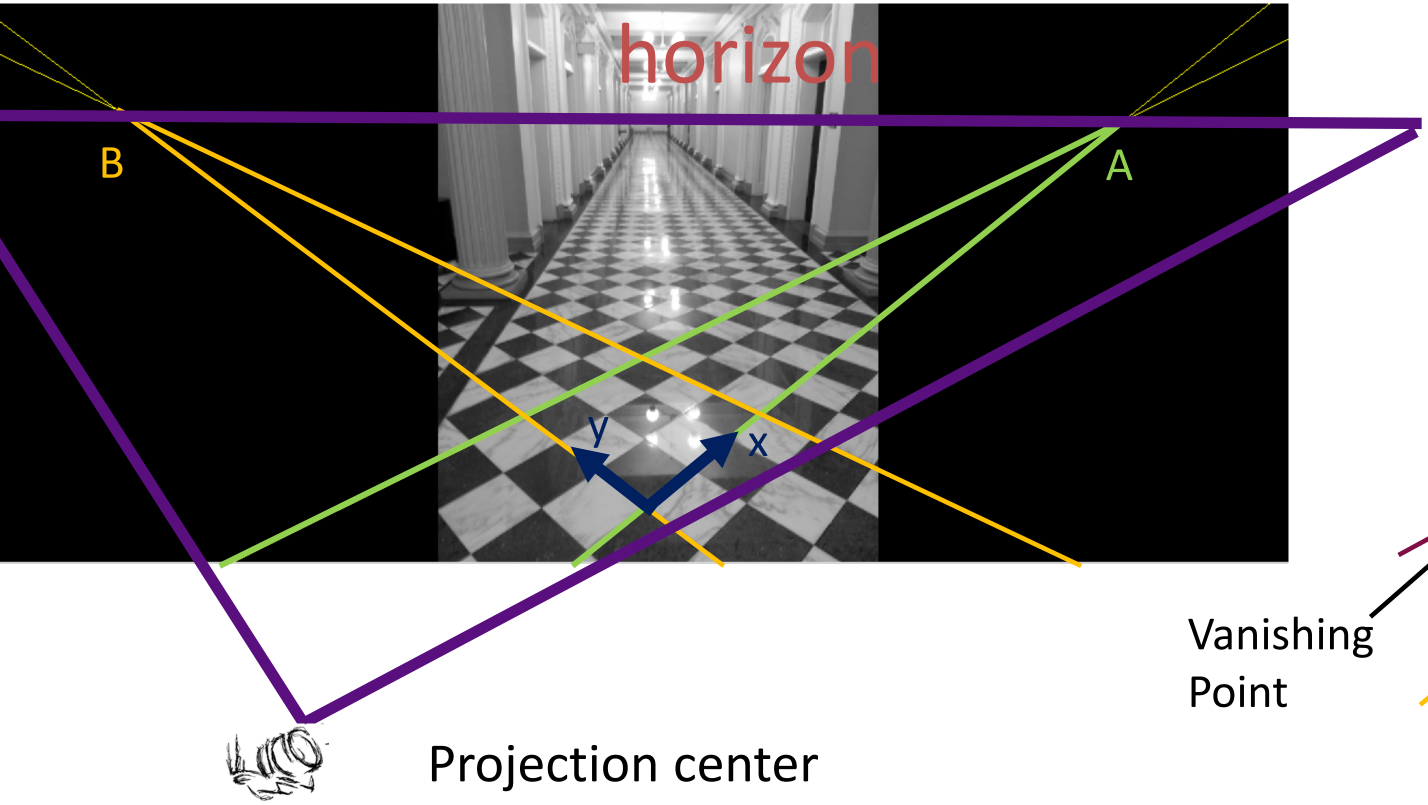
How can we compute the projective transformation between a known pattern and its projection?



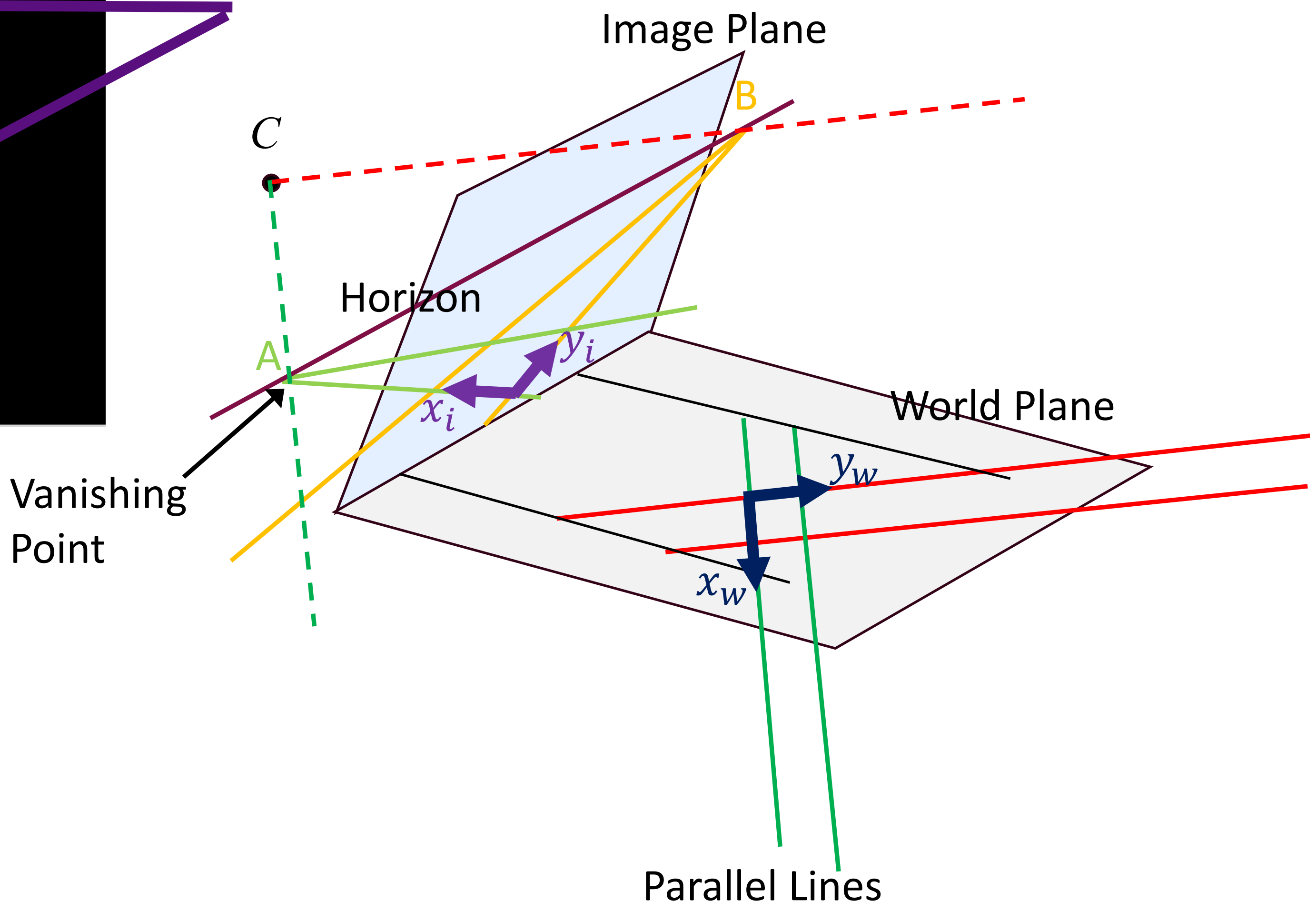


These are homogeneous coordinates to represent the known pattern in $\mathbb{P}^2$

Cont.



Projection center



Assume that a mapping H maps the three points $(1,0,0)$, $(0,1,0)$, and $(0,0,1)$ to the non-collinear points A,B,C

with coordinate vectors a, b and $c \in \mathbb{P}^2$. Then the following is a possible projective transformation:

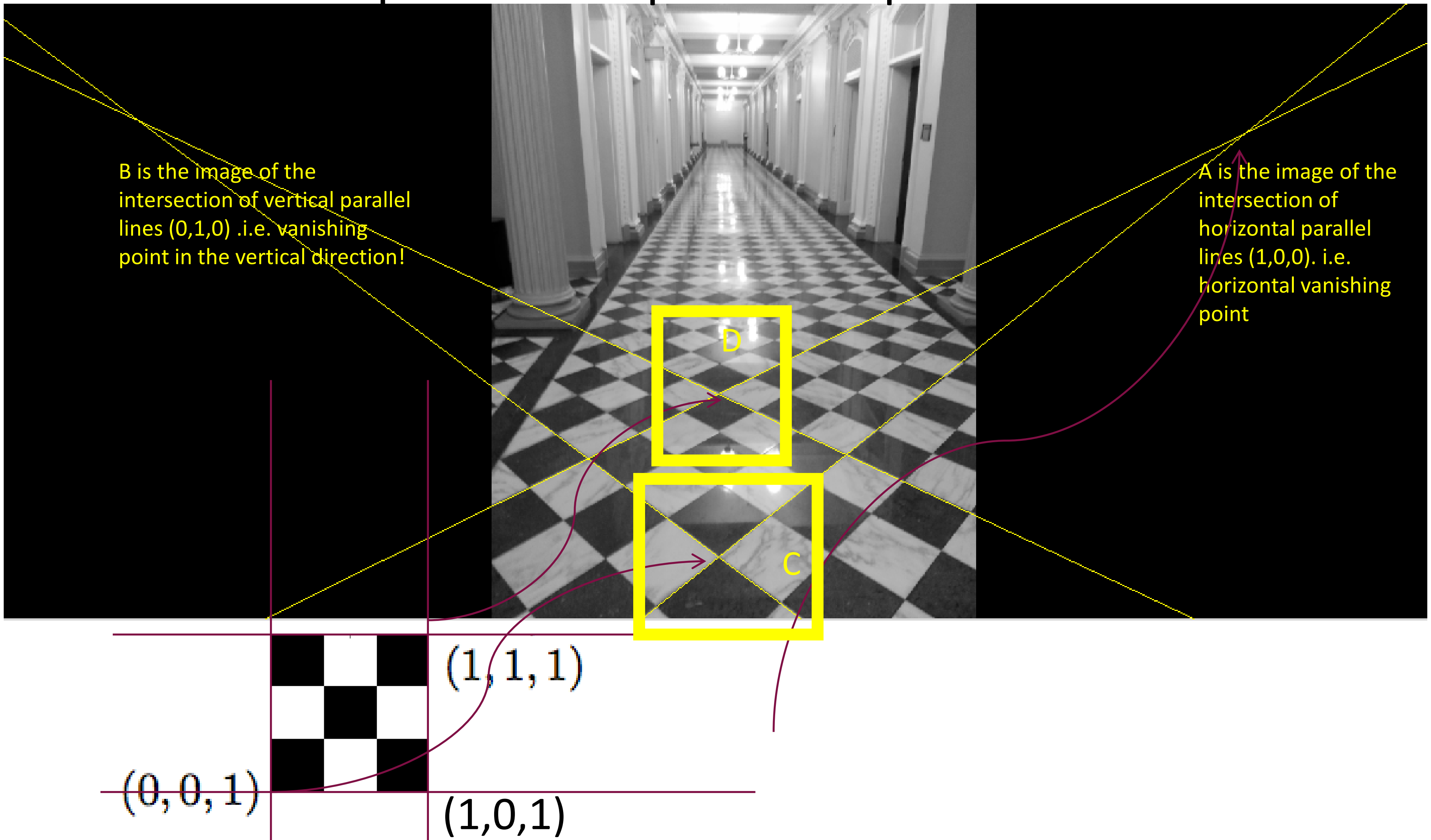
$$\begin{pmatrix} a & b & c \end{pmatrix} = \underbrace{\begin{pmatrix} \alpha a & \beta b & \gamma c \end{pmatrix}}_{H_{3\times 3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

with 3 degrees of freedoms α, β and γ. This means 3 points do not suffice to compute a projective transformation.

Side note: the first 2 columns of the homography are vanishing points!

Solution: Introduce a 4th point correspondence D

Note: makes sense, because after all, H has 8 degrees of freedom, and each 2D point correspondence pins down 2DOF.



Let us assume that the same H maps $(1,1,1)$ to the point d. Then, the following should hold:

$$\lambda d = \begin{pmatrix} H\,a & \beta b & \gamma c \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix},$$

hence

$$\lambda d = \alpha a + \beta b + \gamma c.$$

There always exist such $\lambda, \alpha, \beta, \gamma$ because four elements of $\mathbb{R}^3 \setminus \{(0,0,0)\}$ are always linearly dependent.

Because a, b, c are not collinear, there exist unique $\alpha/\lambda, \beta/\lambda, \gamma/\lambda$ for writing this linear combination.

There always exist such $\lambda, \alpha, \beta, \gamma$ because four elements of $\mathbb{R}^3 \setminus \{(0,0,0)\}$ are always linearly dependent.

Because a, b, c are not collinear, there exist unique $\alpha/\lambda, \beta/\lambda, \gamma/\lambda$ for writing this linear combination.

Since H is the same as H/λ we solve for α, β, γ such that $d = \alpha a + \beta b + \gamma c$, which can be written as a linear system

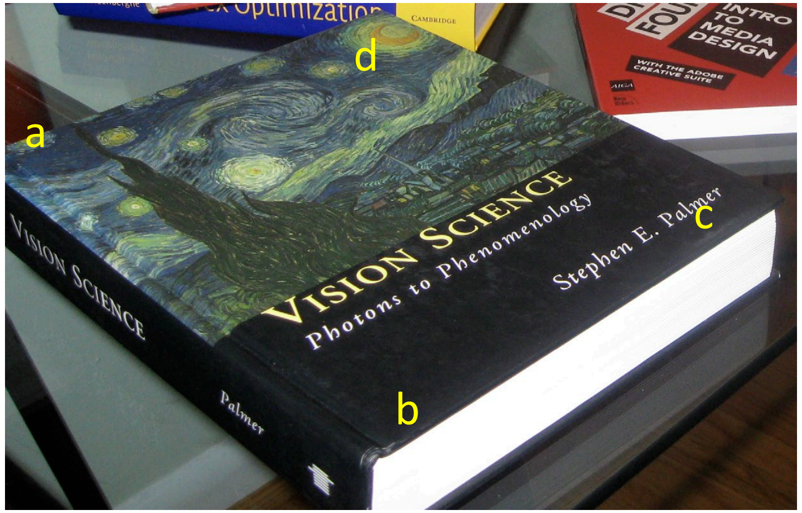
$$\begin{pmatrix} a & b & c \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} = d.$$

Since a, b, c are not collinear we can always find a unique triple α, β, γ. The resulting projective transformation is $H = \begin{pmatrix} \alpha a & \beta b & \gamma c \end{pmatrix}$.

Four points, no three of them collinear, suffice to unambiguously recover a homography

Choosing the points to be the horizontal and vertical vanishing points (1,0,0), (0,1,0) plus origin (0,0,1) and the diagonal (1,1,1) is particularly "nice" especially if you have a square to start from, but really, any four non-collinear points will do.
(coming up next)

What happens when the original set of points is not a square?



c

Find projective transformation mapping $(a, b, c, d) \rightarrow (a', b', c', d')$:

To determine this mapping we go through the four canonical points.

We find the mapping from $(1,0,0), etc$ to (a,b,c,d) and we call it T:

$$a \sim T(1,0,0)^T, etc$$

We find the mapping from $(1,0,0), etc$ to (a',b',c',d') and we call it T':

$$a' \sim T'(1,0,0)^T, etc$$

Then, back-substituing $(1,0,0)^T \sim T^{-1}a, etc$ we obtain that

$$a' = T'T^{-1}a, etc$$

To determine this mapping we go through the four canonical points.

We find the mapping from $(1,0,0), etc$ to (a,b,c,d) and we call it T:

$$a \sim T(1,0,0)^T, etc$$

We find the mapping from $(1,0,0), etc$ to (a',b',c',d') and we call it T':

$$a' \sim T'(1,0,0)^T, etc$$

Then, back-substituing $(1,0,0)^T \sim T^{-1}a, etc$ we obtain that

$$a' = T'T^{-1}a, etc$$

To determine this mapping we go through the four canonical points.

We find the mapping from $(1,0,0), etc$ to (a,b,c,d) and we call it T:

$$a \sim T(1,0,0)^T, etc$$

We find the mapping from $(1,0,0), etc$ to (a',b',c',d') and we call it T':

$$a' \sim T'(1,0,0)^T, etc$$

Then, back-substituing $(1,0,0)^T \sim T^{-1}a, etc$ we obtain that

$$a' = T'T^{-1}a, etc$$

This amounts to computing homographies from a made-up square, then inverting one of them and composing!