CIS 5800

Machine Perception

Instructor: Lingjie Liu

Lec 7: Feb 17, 2025

Administrivia

- Additional PreHW TA Office Hour:

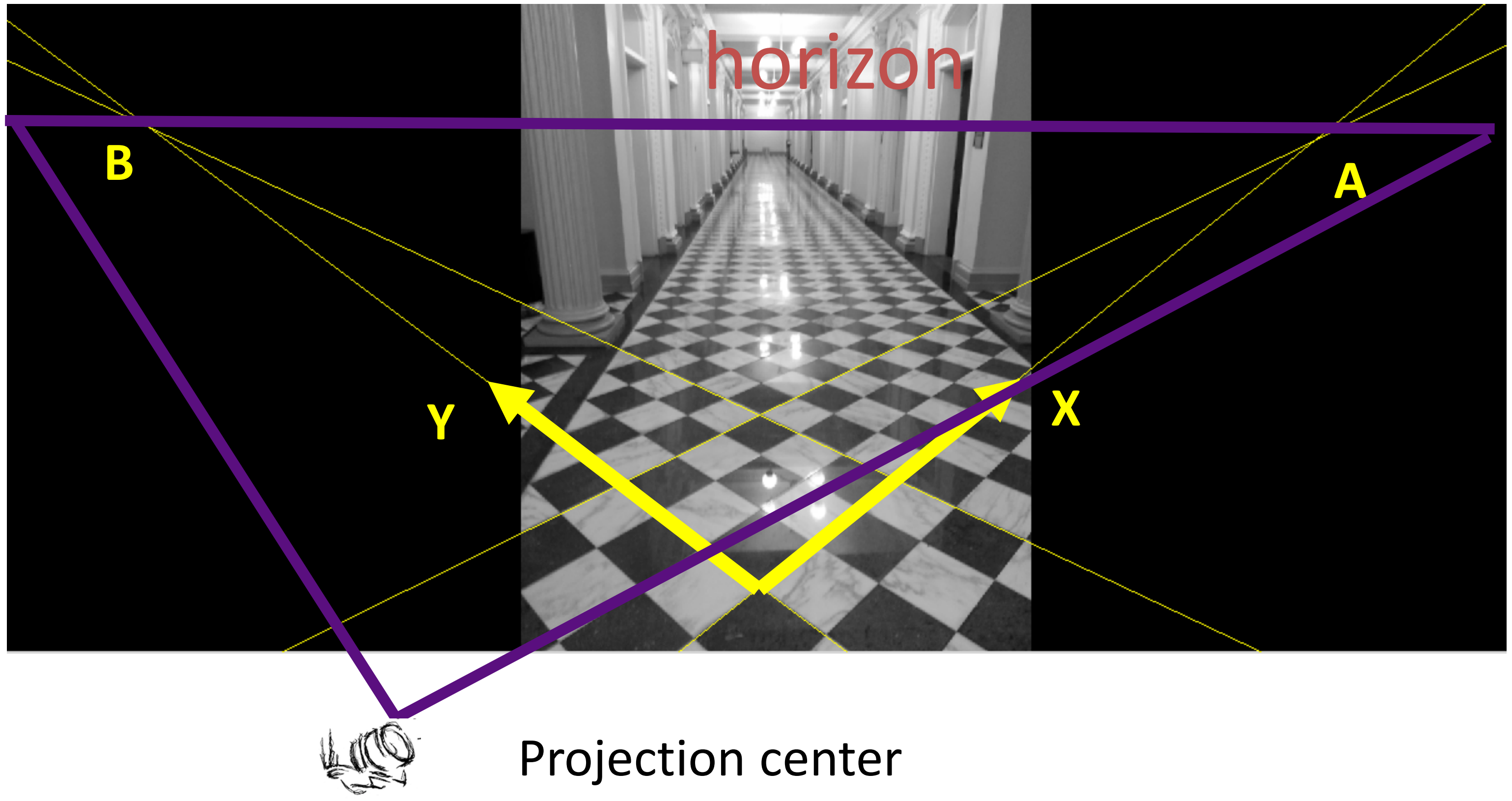
Yiming Huang: Tuesday Feb 18 PM -1PM: https://upenn.zoom.us/j/8013153196

- HW1 Due on next Wednesday 11:59pm ET.

Recap:

Vanishing rays/planes through the camera center are parallel to the world lines/planes

So, the horizon plane is parallel to the ground plane and hence $h_1 \times h_2$ is the normal to the ground plane!



Horizon plane = Vanishing plane = Viewing plane

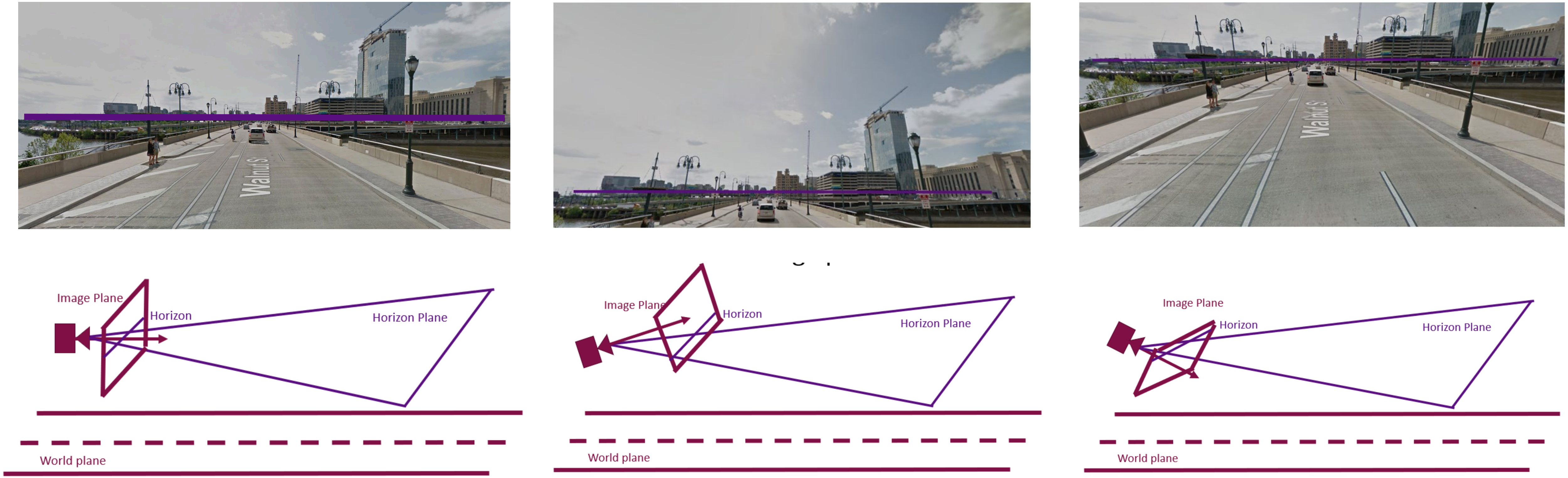
World plane // vanishing plane

World plane = Ground plane in this case

Recap:

Horizon gives complete info about how camera is oriented w.r.t. world plane*!

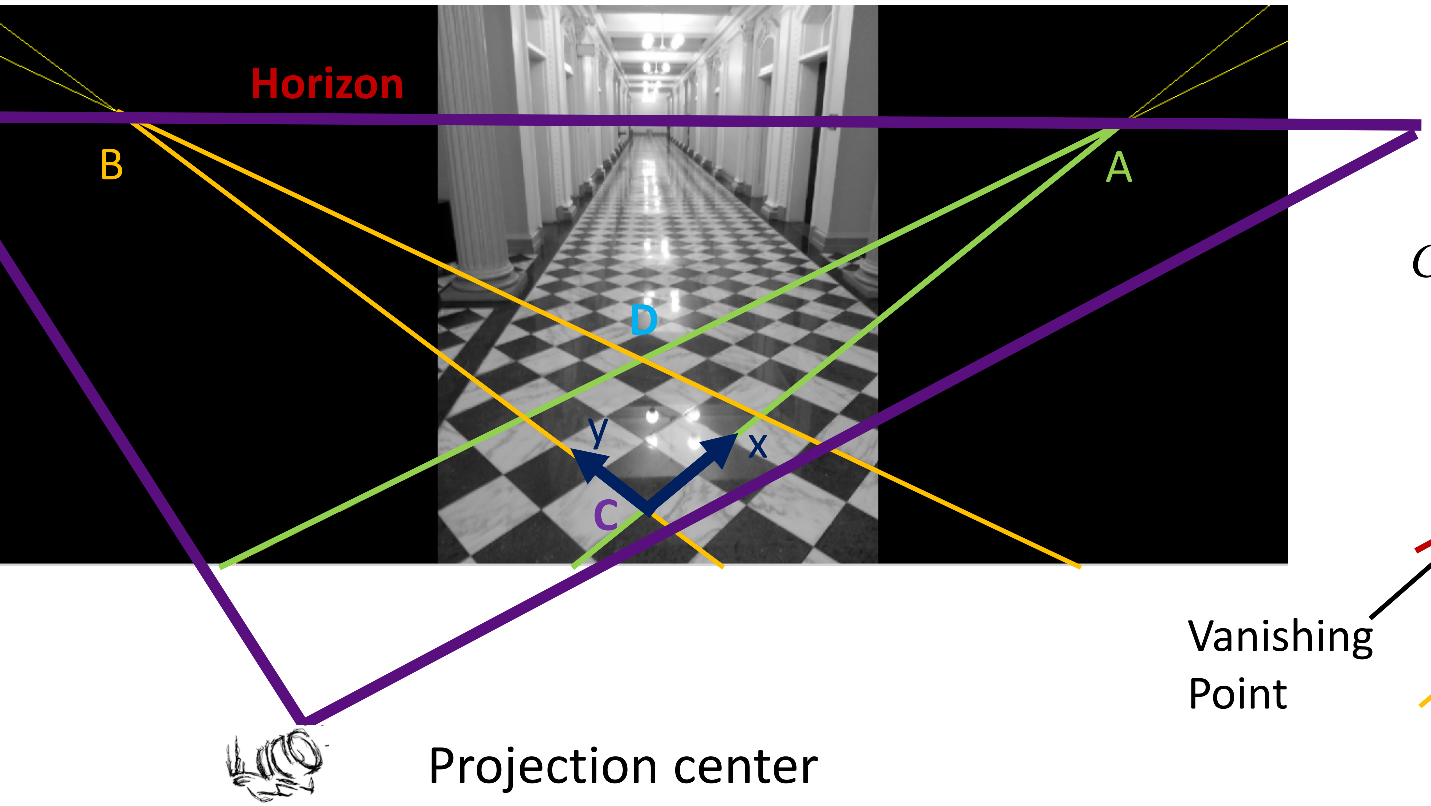
Thumb rule: “If horizon is horizontal & central, camera is correctly vertical & principal axis is parallel to world plane**!”

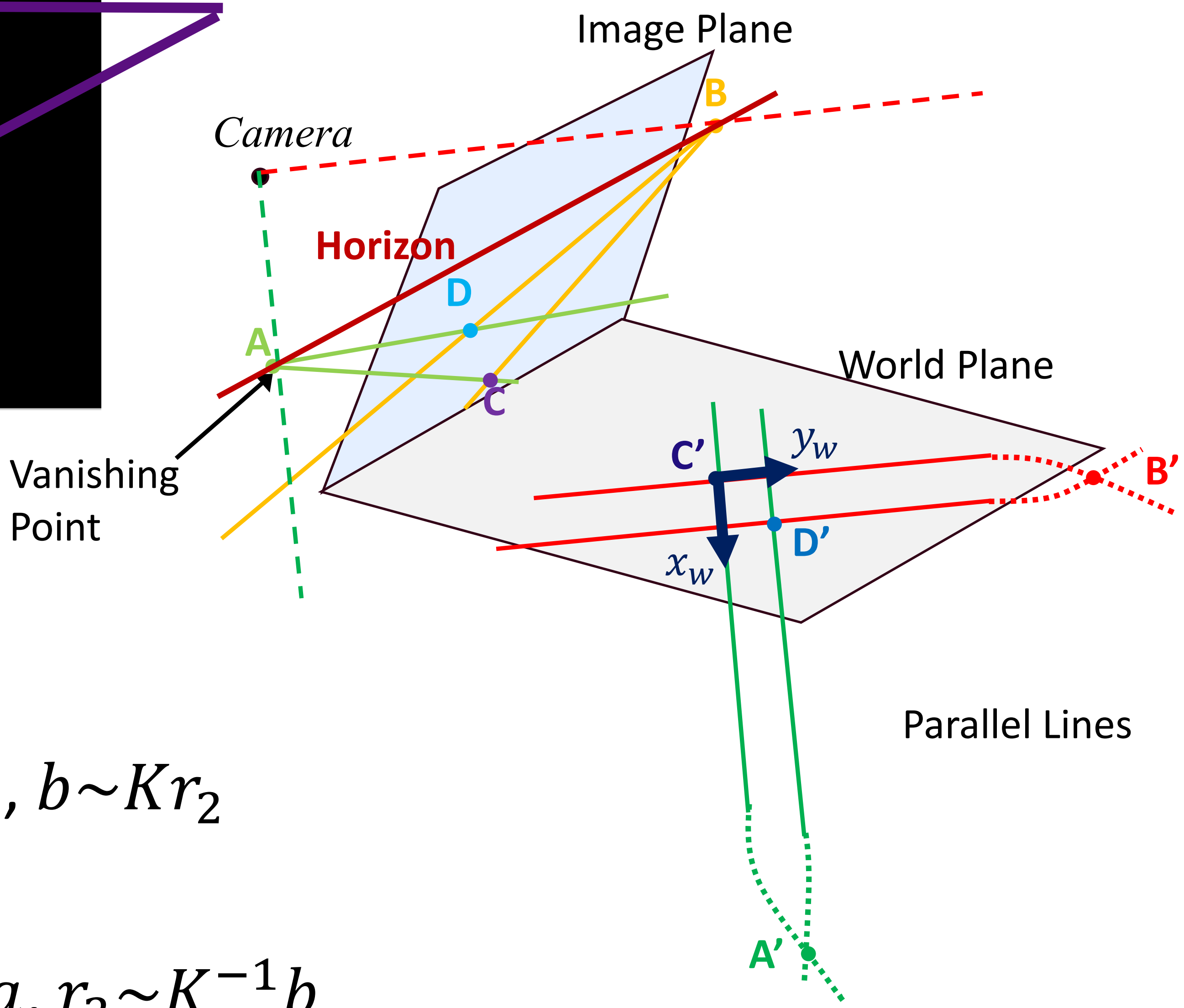


*caveat: assuming known K

**caveat: assuming that principal axis passes through image center, and camera axes are horizontal. (usually approximately true)

Recap:





$$H \sim K(r_1, r_2, T)$$

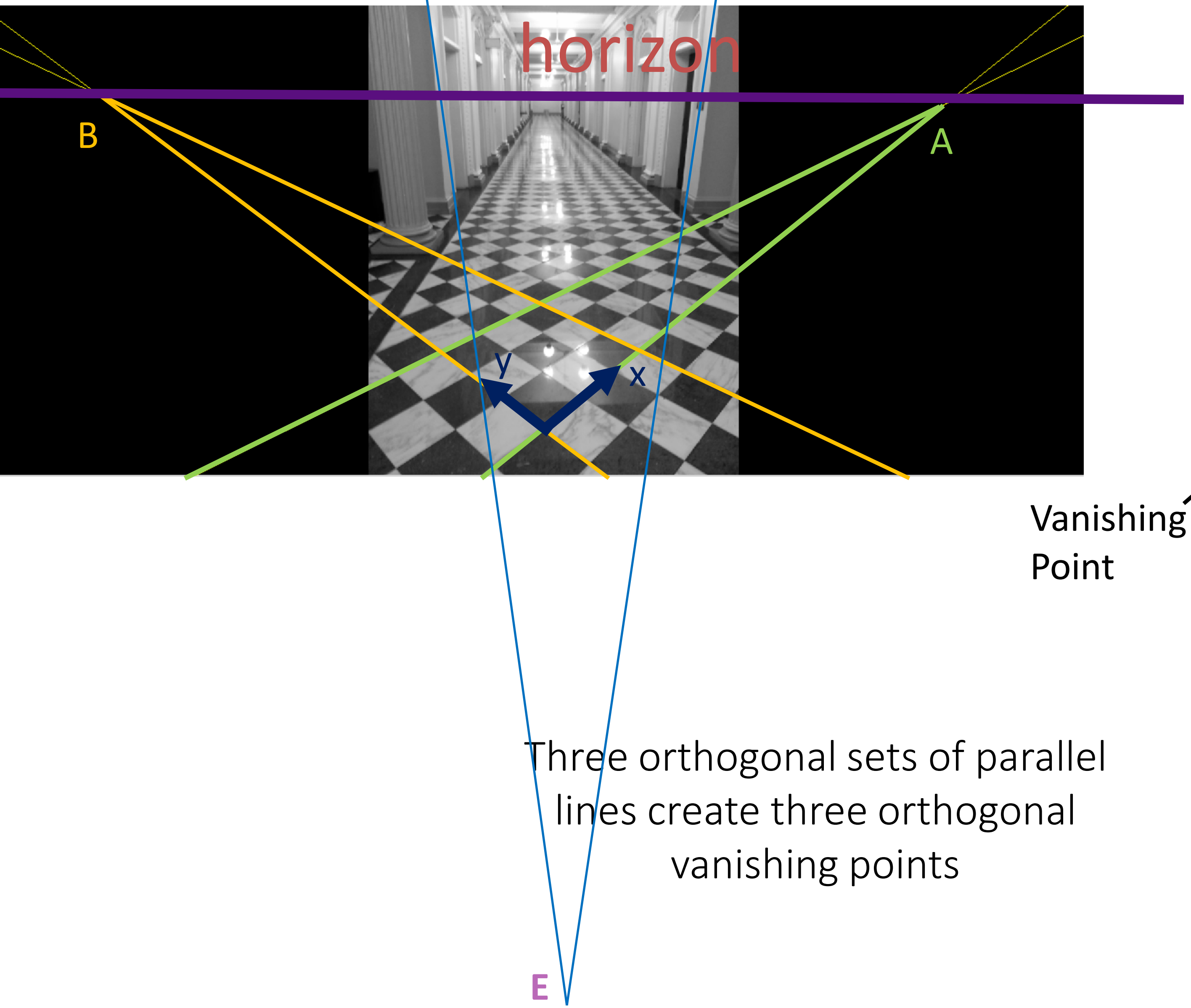
$$H \sim (a, b, c)$$

$$a \sim K r_1,\ b \sim K r_2$$

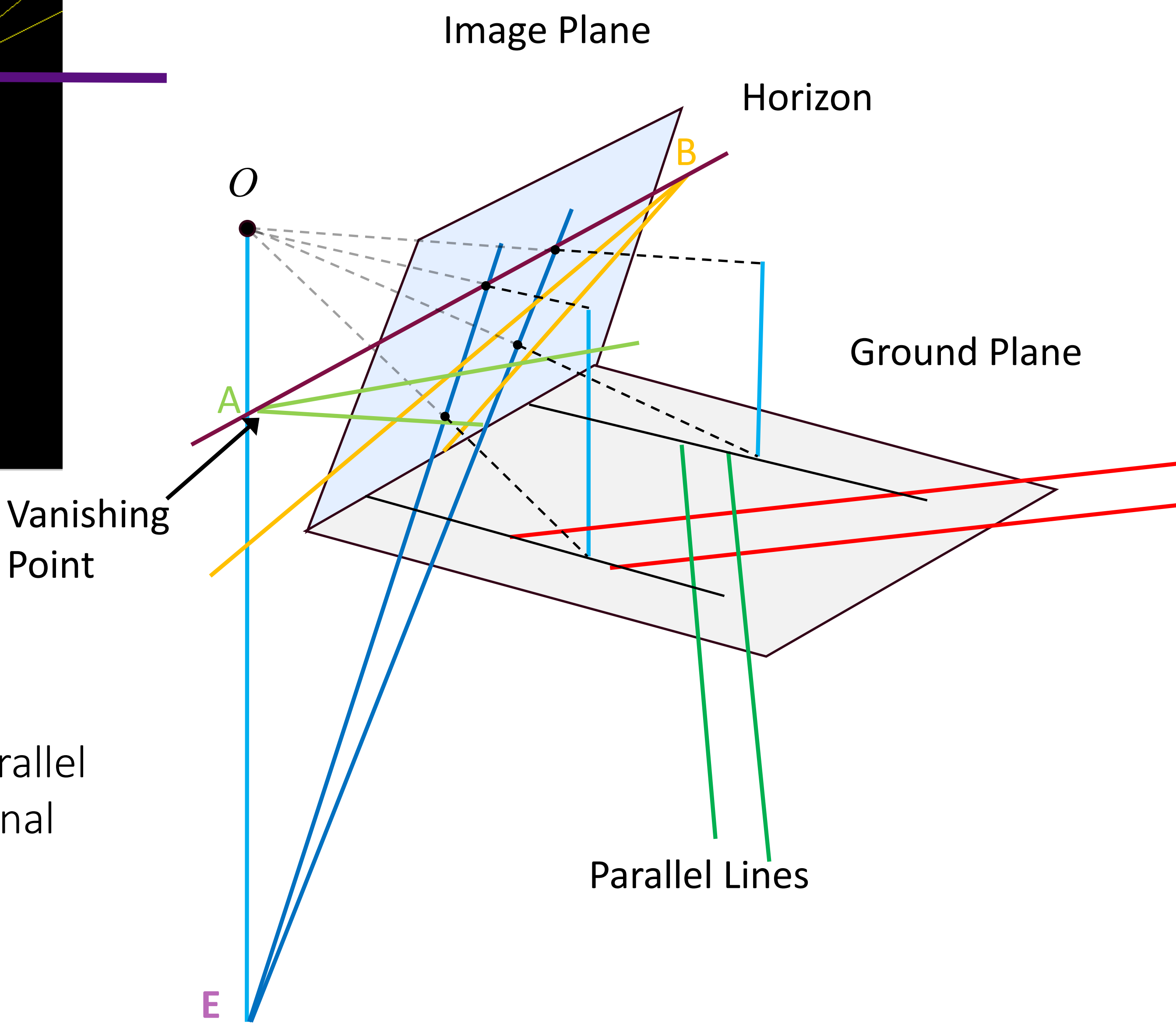
$$r_1 \sim K^{-1} a,\ r_2 \sim K^{-1} b$$

Recap:

A scene with three orthogonal sets of parallel lines

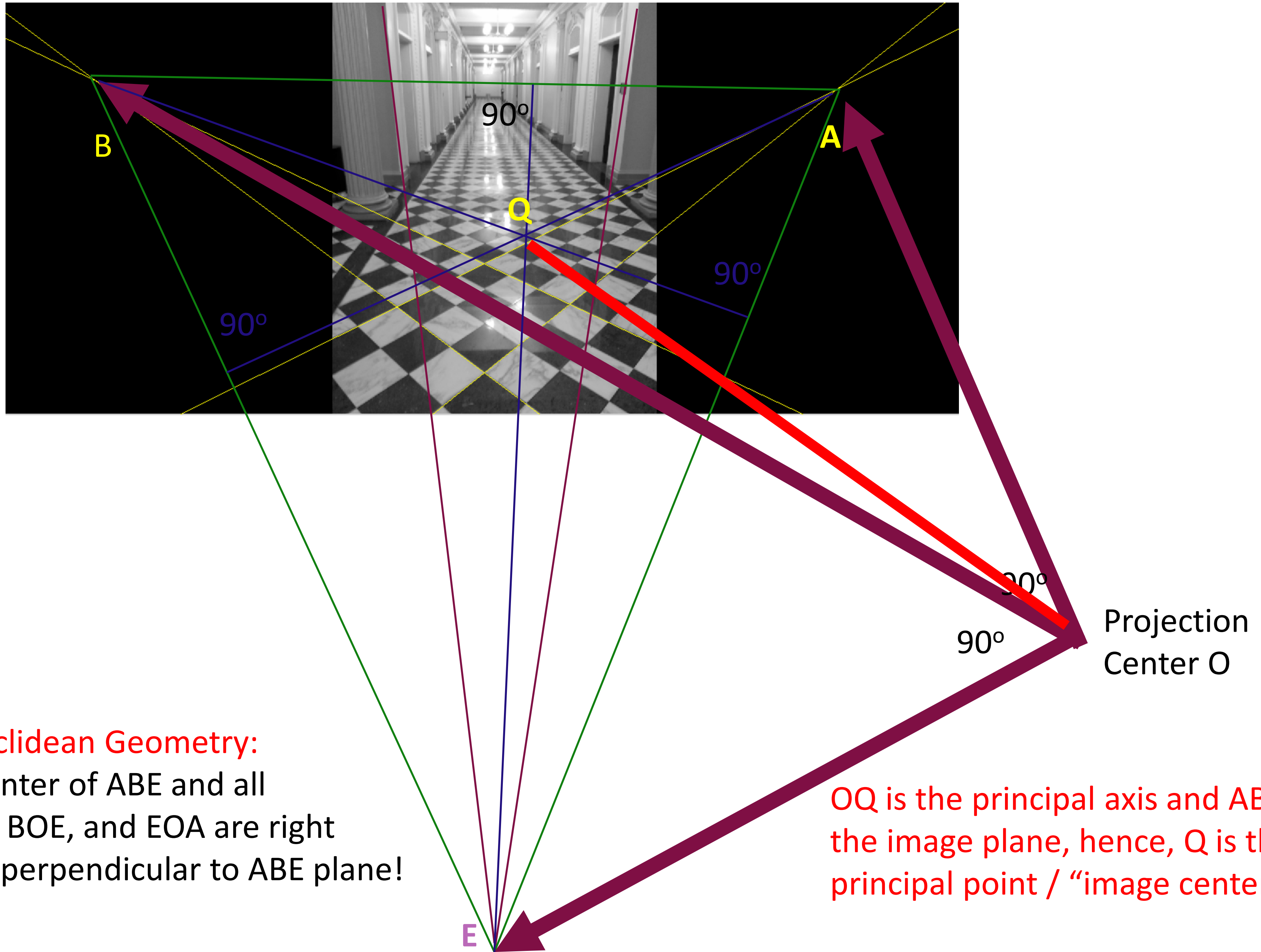


Three orthogonal sets of parallel lines create three orthogonal vanishing points



Recap:

Let Q be the orthocenter of the triangle ABC



Theorem from Euclidean Geometry:

If Q is the orthocenter of ABE and all three angles AOB, BOE, and EOA are right angles , the OQ is perpendicular to ABE plane!

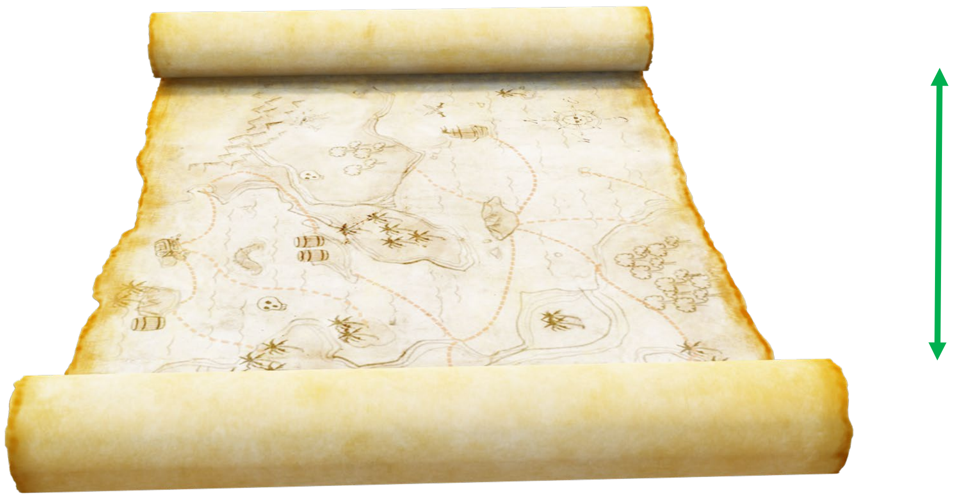
OQ is the principal axis and ABE is the image plane, hence, Q is the principal point / “image center”

Recap: Summary

- If we have 3 orthogonal VPs, we can get full intrinsics K (focal length and image center), and also extrinsics R.
  - What’s missing? Just the translation t.
    - And that information is not contained in VPs, because camera translations don’t affect the VPs!
    - Which is why, when we found homographies (that do contain full information about translation), we used more than just VPs.

Cross Ratios & Length Measurements from Single Images ("Single View Metrology")

Are lengths preserved under homography?

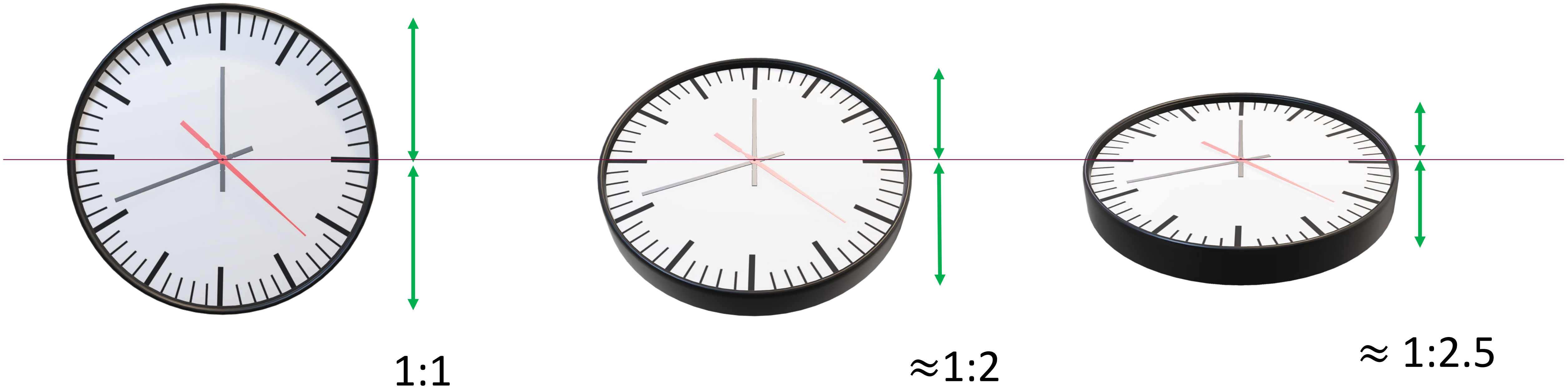


Obviously not.

What about length ratios?

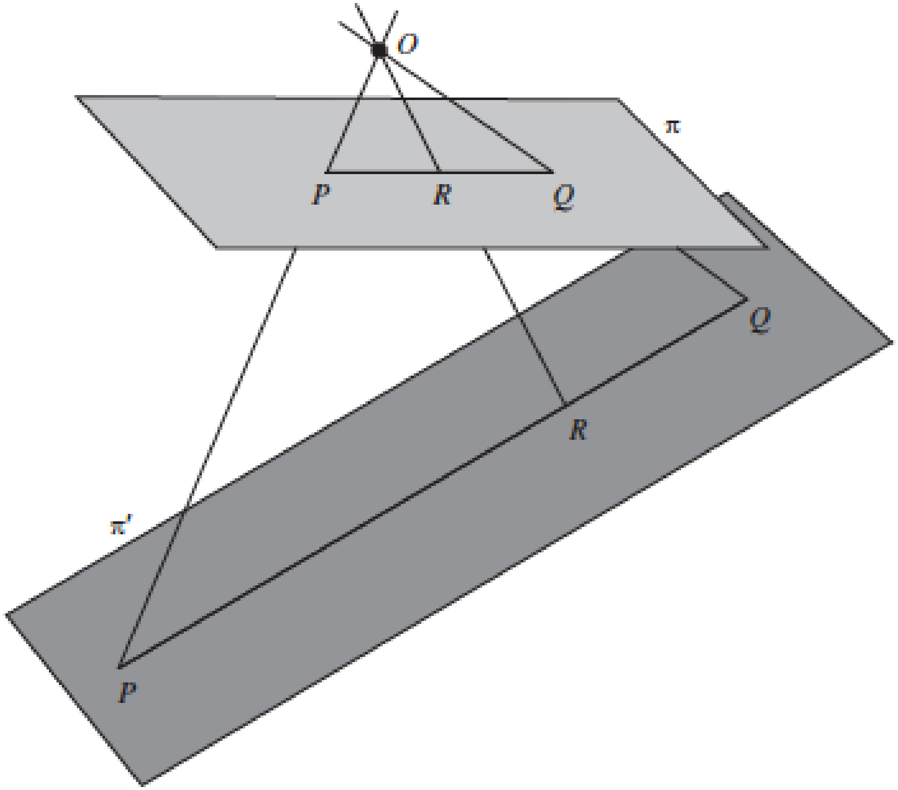
Projection of a circle

From other perspective views, would the center of this circular clock face remain at center?



Clearly, length ratios are not preserved under homographies!

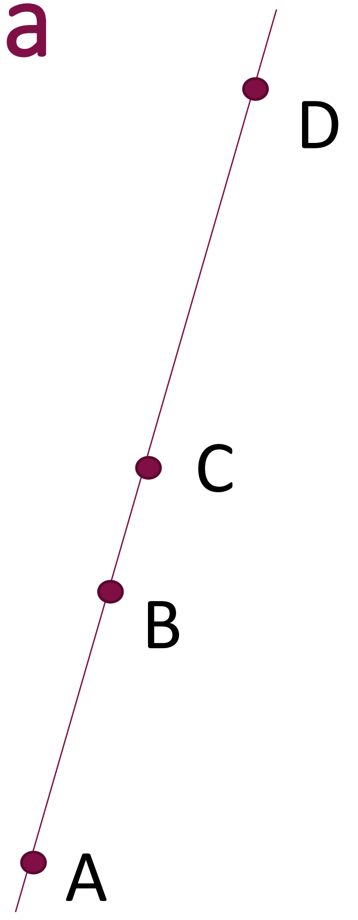
Length ratios under homography



Clearly, length ratios are not preserved under homographies!

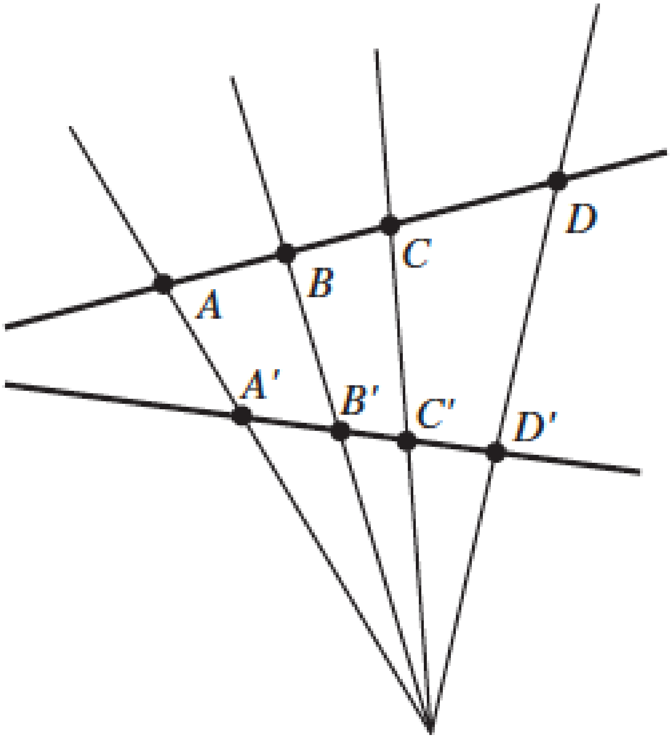
What metric property along a line *is preserved* under a projective transformation?

- Not lengths AB i.e. distance of two pts from each other.
- Not length ratios i.e. distance of two pts from a third collinear pt. AC: BC



- Instead, what is preserved is:
  - **Ratio of ratios** of distances of two pts from two other collinear pts.
    - $\frac{AC}{AD} : \frac{BC}{BD}$ = Cross ratio of A, B, C, D
- CR can also be written as $\frac{AC.BD}{AD.BC}$

Cross Ratios of Collinear Points

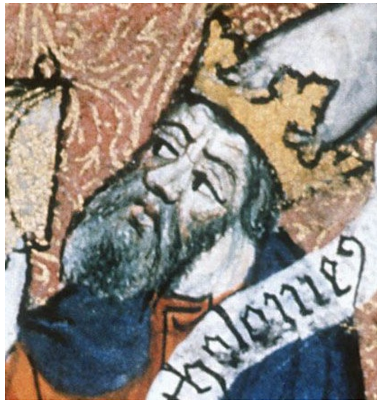


Given four points A, B, C, D, we define the cross-ratio of their distances as

$$CR(A,B,C,D) = \frac{AC}{AD} : \frac{BC}{BD}.$$

$CR(A,B,C,D)$ remains invariant under projective transformations

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



Pappus (290-350 AD)

Denote $\mathbf{x}'_i = (x'_i, 1)^T$ and $\mathbf{x}_i = (x_i, 1)^T$. Suppose $\mathbf{x'}_i = H\mathbf{x}_i$. Note $|H| = \det(H)$. Then

$$x'_i - x'_j = |\mathbf{x}'_i \;\; \mathbf{x}'_j| = |\lambda_i H\mathbf{x}_i \;\; \lambda_j H\mathbf{x}_j| = \lambda_i\lambda_j |H(\mathbf{x}_i \;\; \mathbf{x}_j)|$$

$$= \lambda_i\lambda_j |H||\mathbf{x}_i \;\; \mathbf{x}_j| = \lambda_i\lambda_j \det(H)(x_i - x_j)$$

Note: C

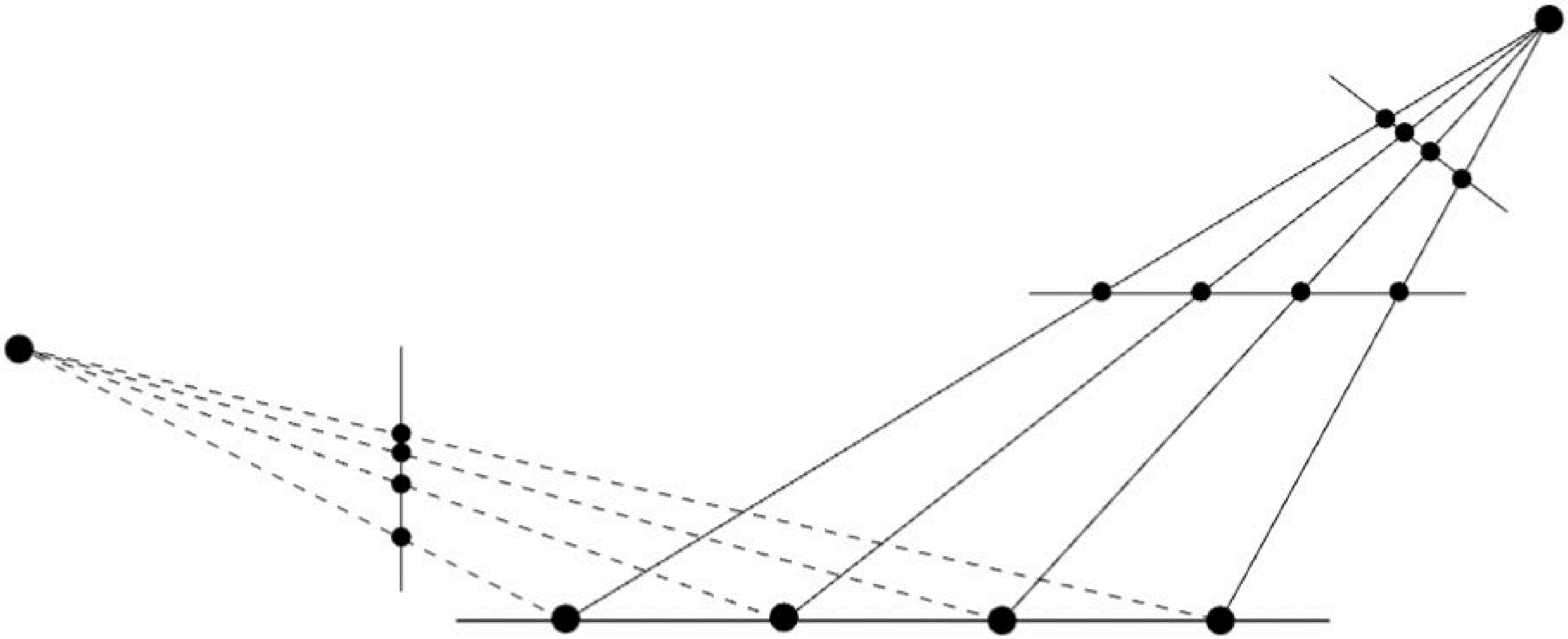
but it is

e.g., CR

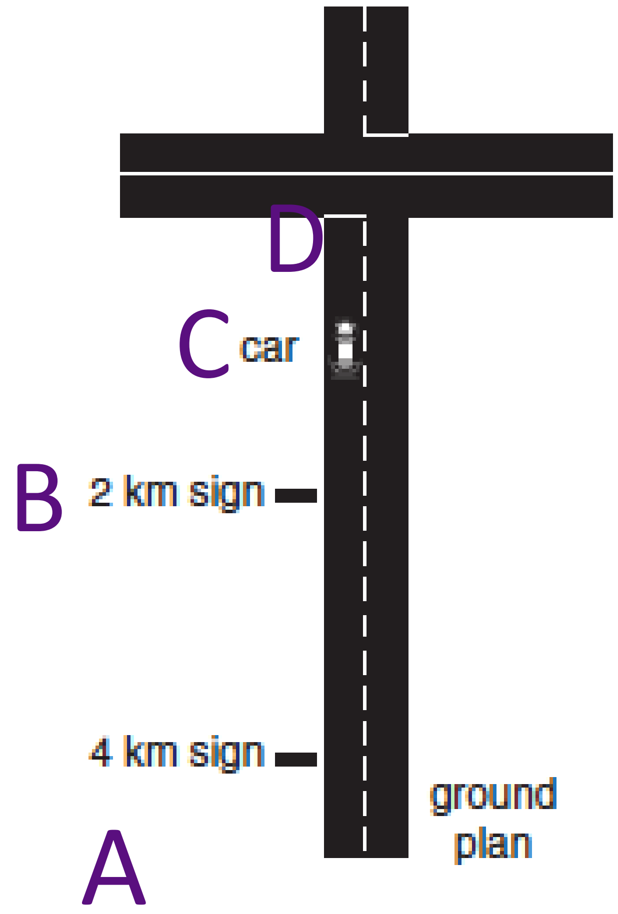
Brannan et a

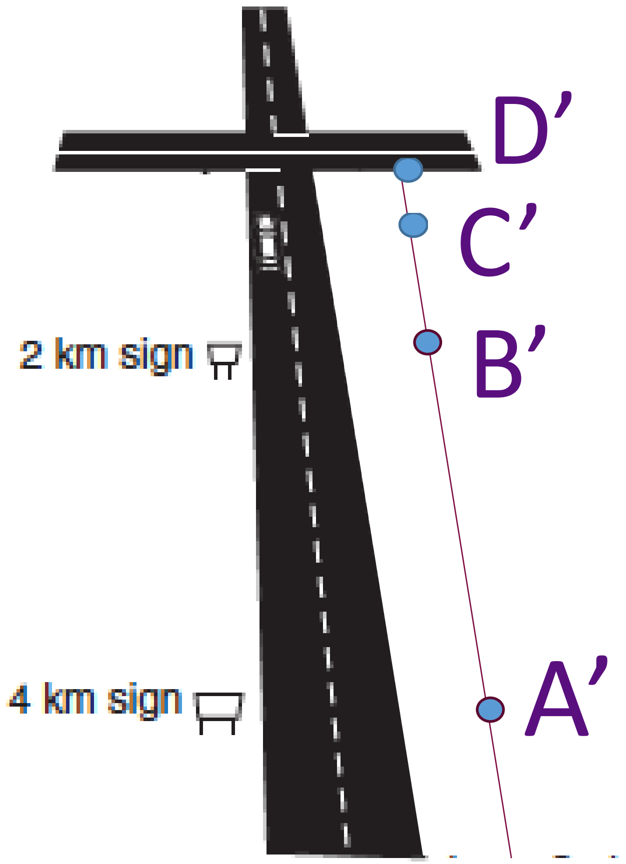

Girard Desargues (1591-1661)

Same cross ratio for all these 4 point sets!



Example: Cross ratios for metrology





How far is car from intersection:

CD=?

A’D’=300 pixels

A’C’=275pixels

B’C’=50 pixels

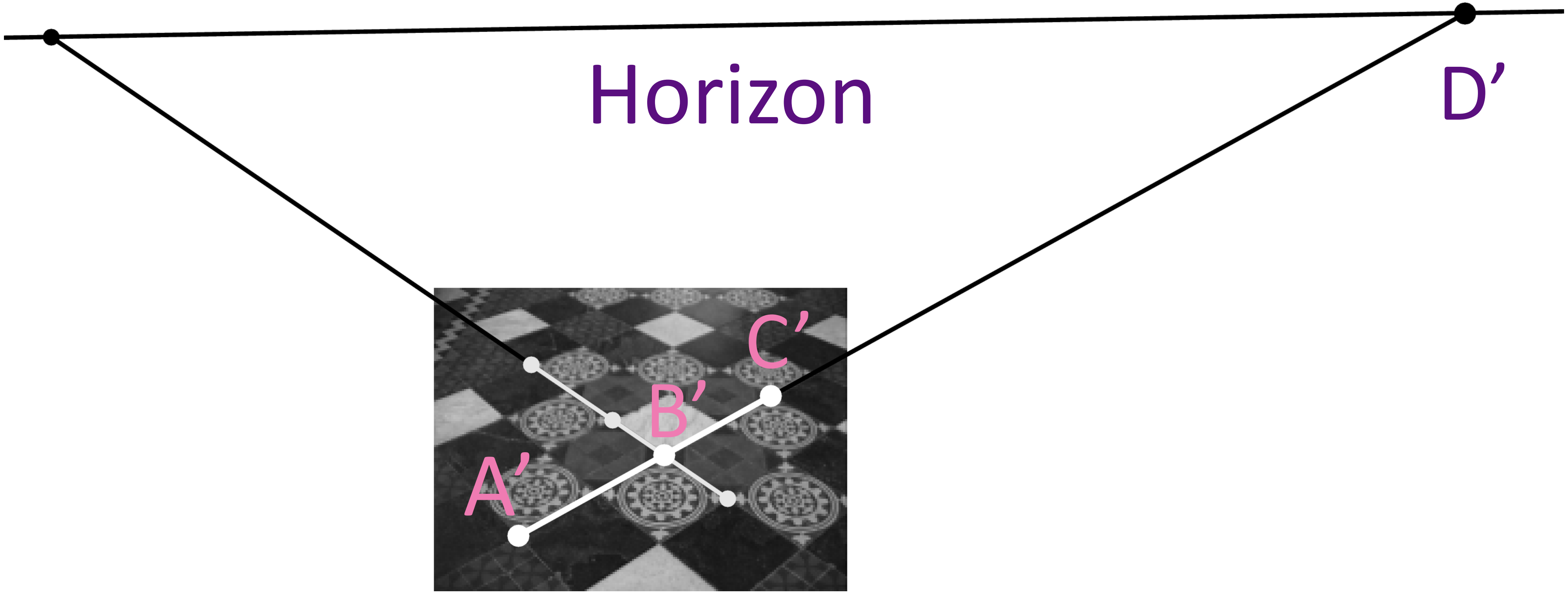
$$\frac{A'C'}{A'D'} : \frac{B'C'}{B'D'} = \frac{275}{300} : \frac{50}{75} = 1.375$$

$$\frac{AC}{AD} : \frac{BC}{BD} = \frac{4-CD}{4} : \frac{2-CD}{2} = 1.375$$

$$CD = 0.857km$$

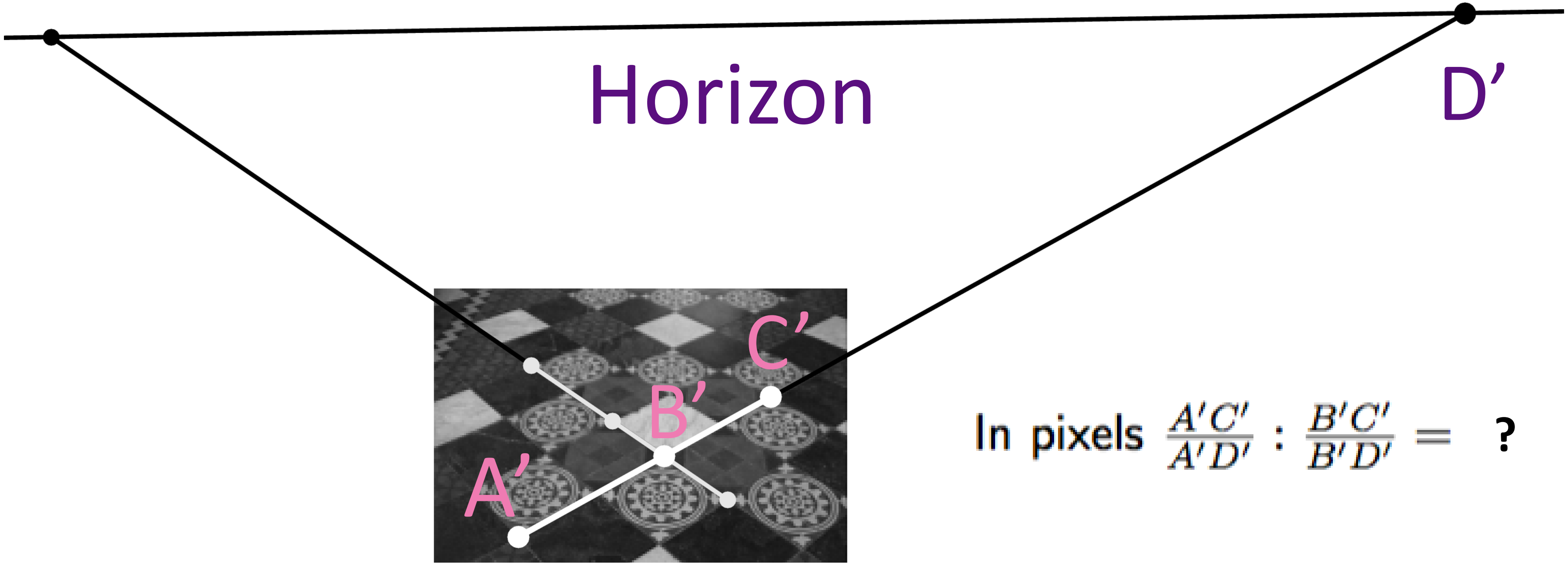
Note: all distances in image are measurable. And 2 distances in world are given.

What happens when one of the points is at infinity?



While D’ is a finite point, D on the original plane is at infinity !

What happens when one of the points is at infinity?

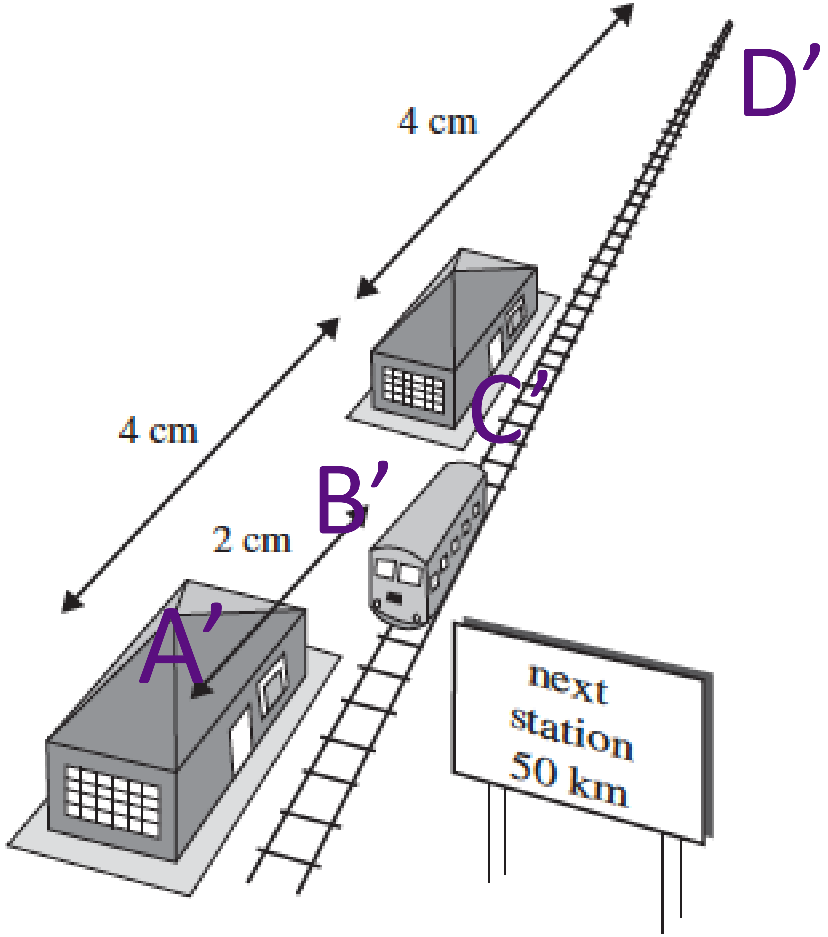


In pixels $\frac{A'C'}{A'D'} : \frac{B'C'}{B'D'} =$?

When a point D is at infinity, the cross-ratio becomes a ratio !

$\frac{AC}{AD} : \frac{BC}{BD} = \frac{AC}{BC}$ (Think $\frac{AC}{\infty} : \frac{BC}{\infty} = \frac{AC}{BC} \times \frac{\infty}{\infty} = \frac{AC}{BC}$)

Vanishing points allow us to measure length ratios!



How far away is the train from the next station? i.e. Length of BC?

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

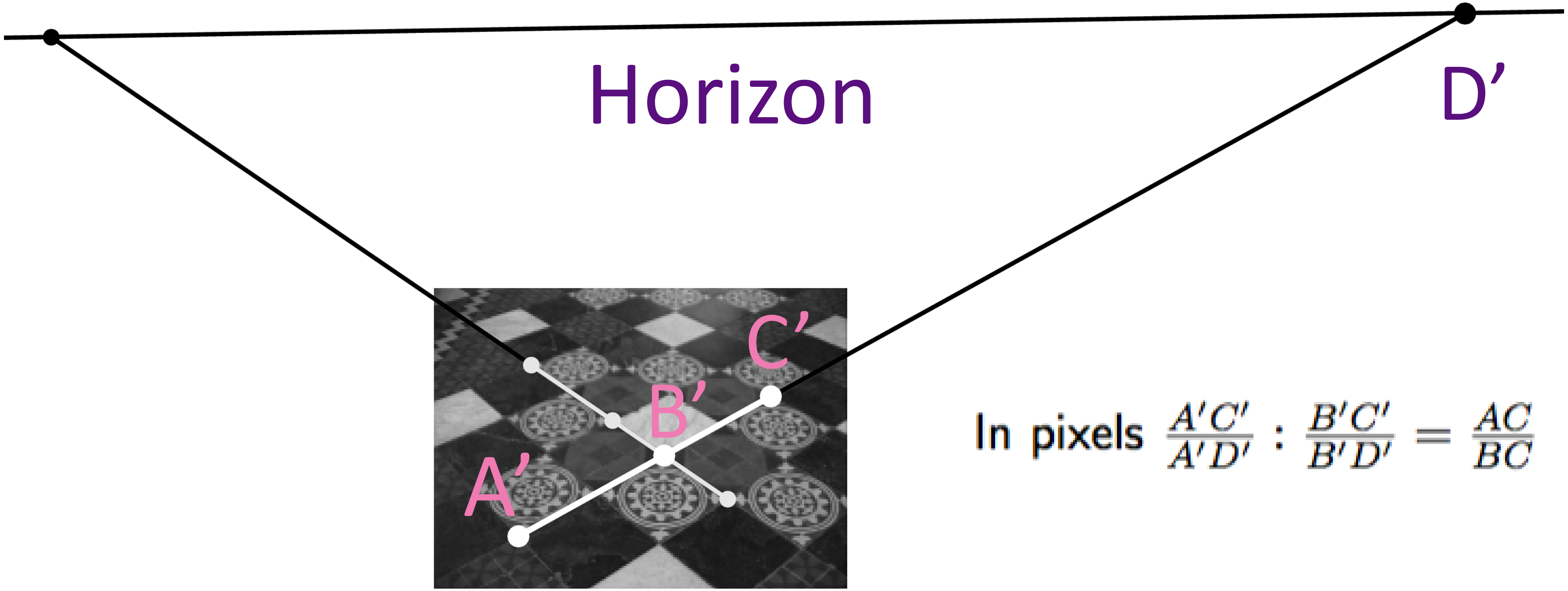
$$\frac{4}{8} : \frac{2}{6} = \frac{50}{BC}$$

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

Note: We only needed measurement of 1 distance in the real world.

Brannan et al. Geometry

Also, world plane length ratios determine vanishing points!



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

Given: AB=BC

$$\frac{A'C'}{A'D'} : \frac{B'C'}{B'D'} = \frac{AC}{BC} = 2$$

If we know A’, B’, C’ in pixels, we can find D’

In this way, we can find vanishing points and the horizon without even needing parallel lines!

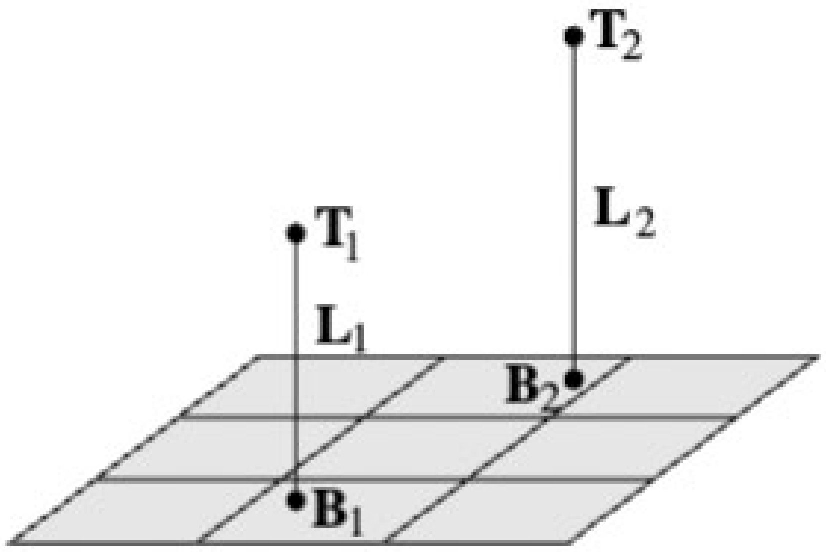
Measuring heights, i.e. distances *from* a world plane

- So far, we've been looking at distances *on* a world plane.

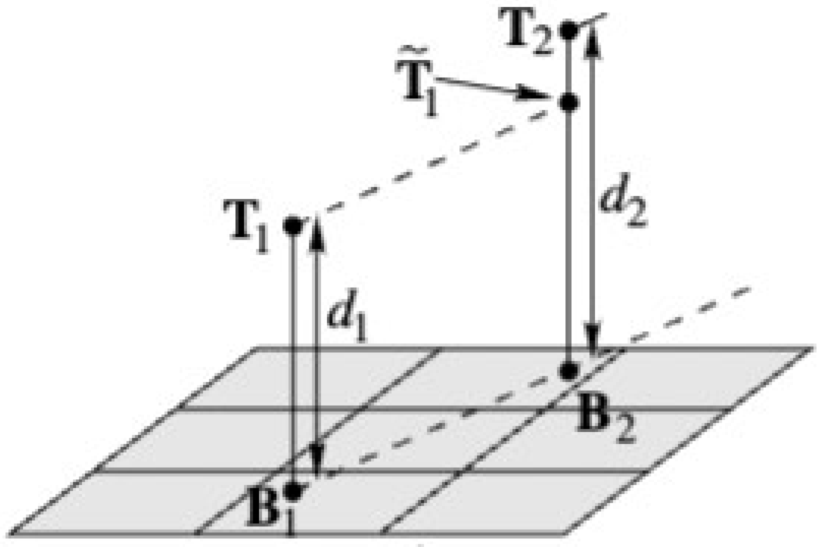
- Next, distances *off it?*

Length transfer in 3D

- In the real 3-D world, you can compare one object with known length to another to "transfer" its length. This is what you do with a ruler, for example.



a



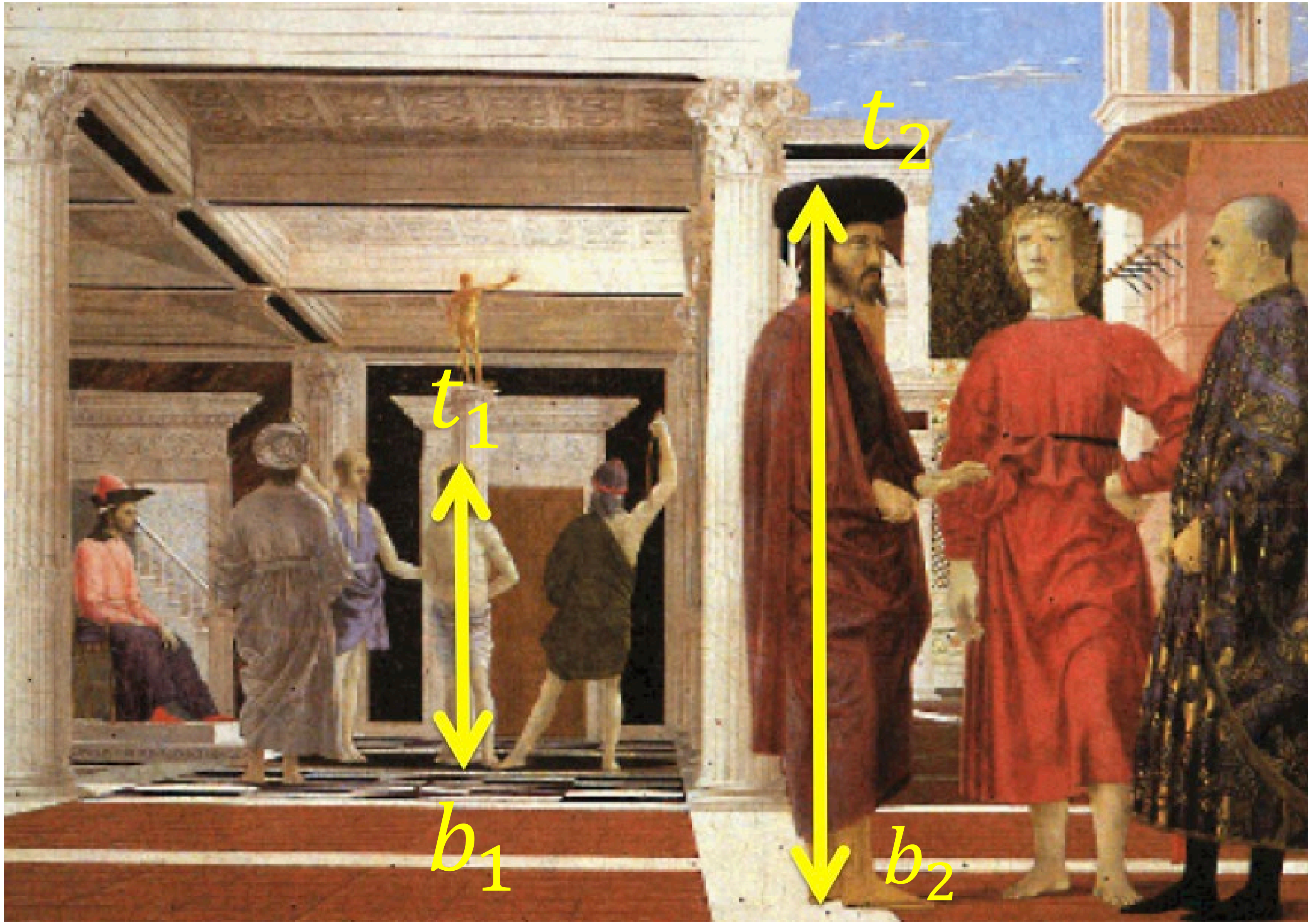
b

How to do this in an image?

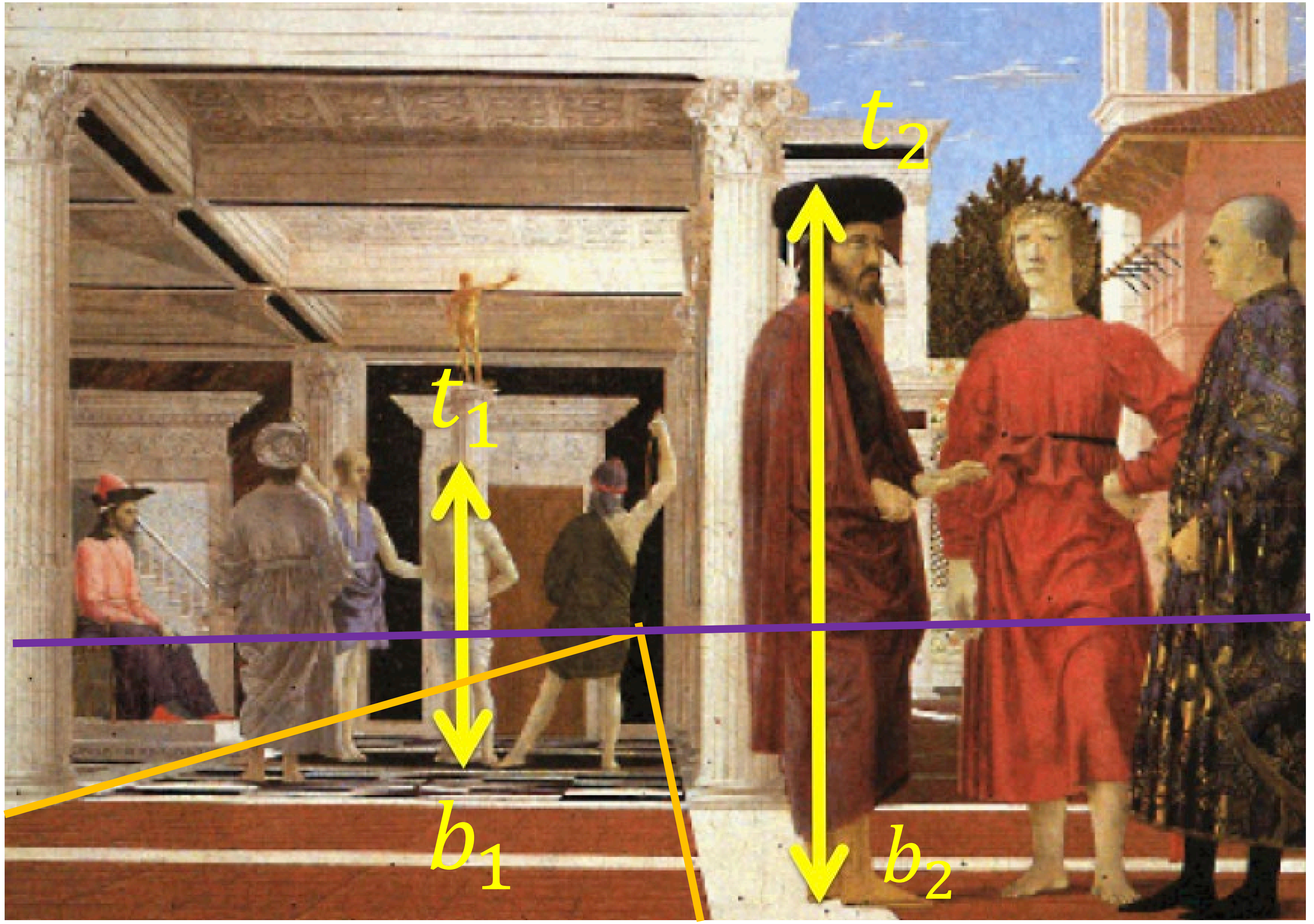
ZH Sec 8.7

Distance Transfer: How tall is the man if the statue is 180cm (in 3D world space)?

Note: we will be going back and forth between world points (denoted as capital letters) and their projections (denoted as small letters).

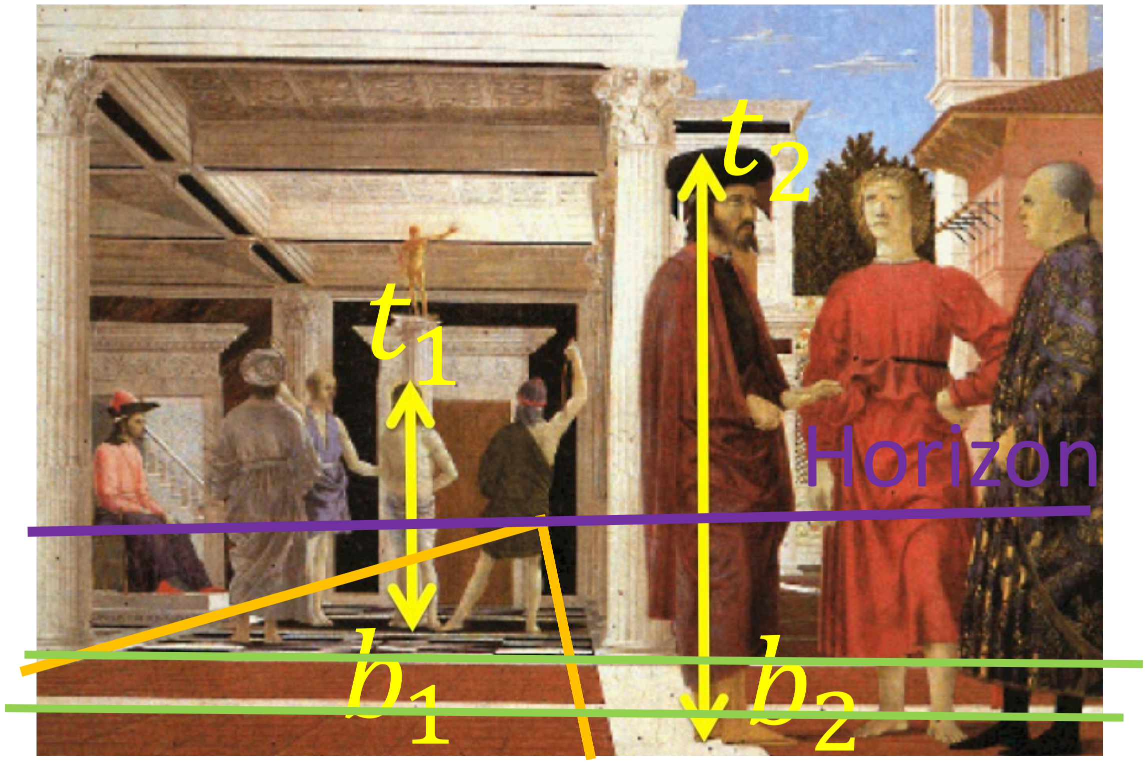


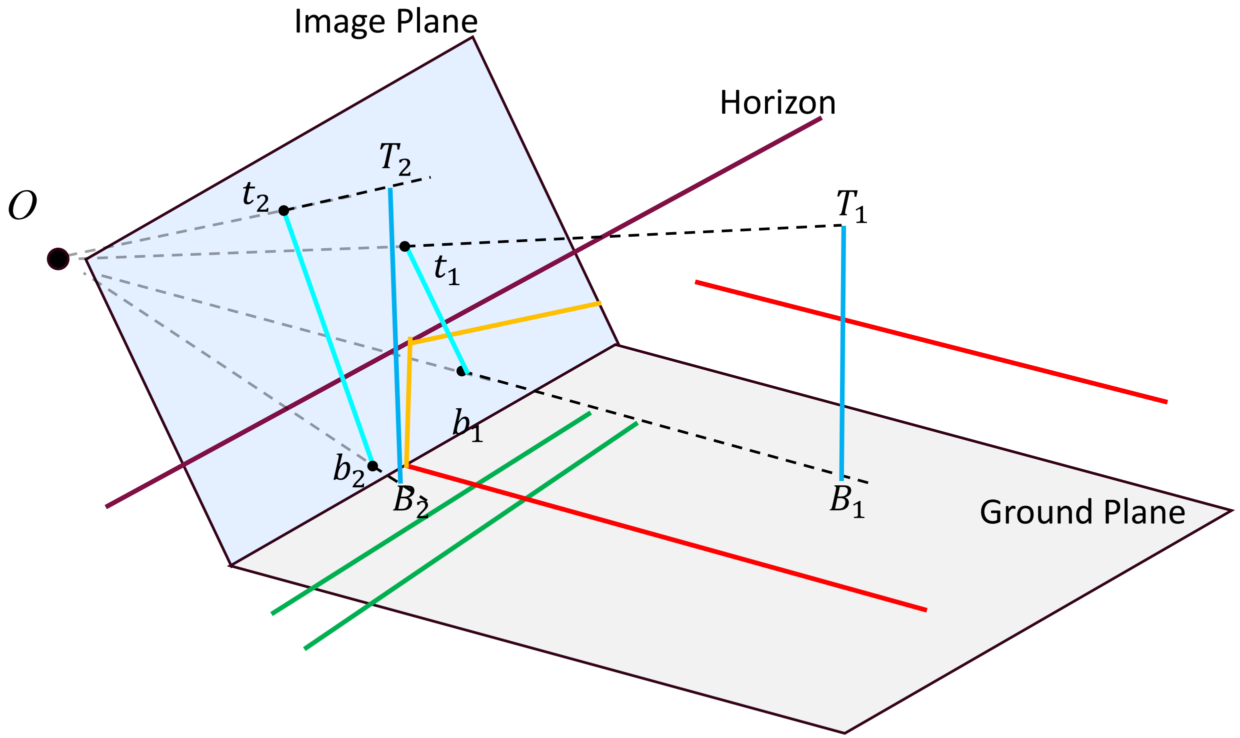
1. Find horizon (from homography, or by finding VPs)



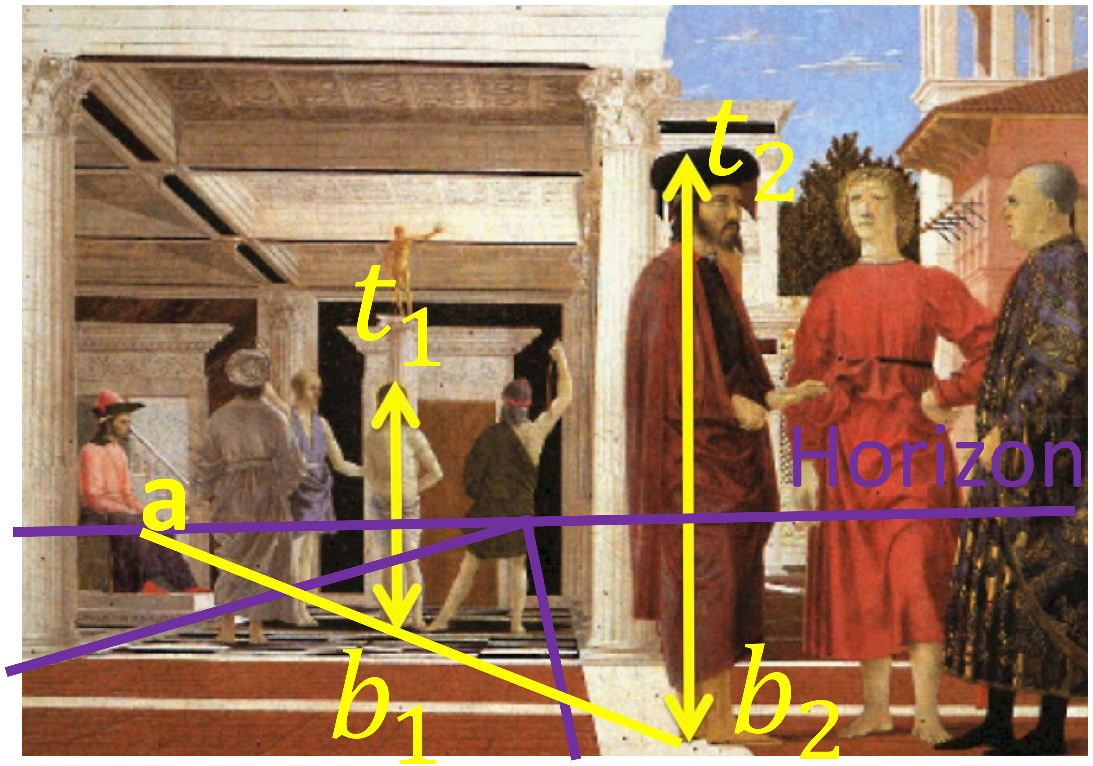
Horizon

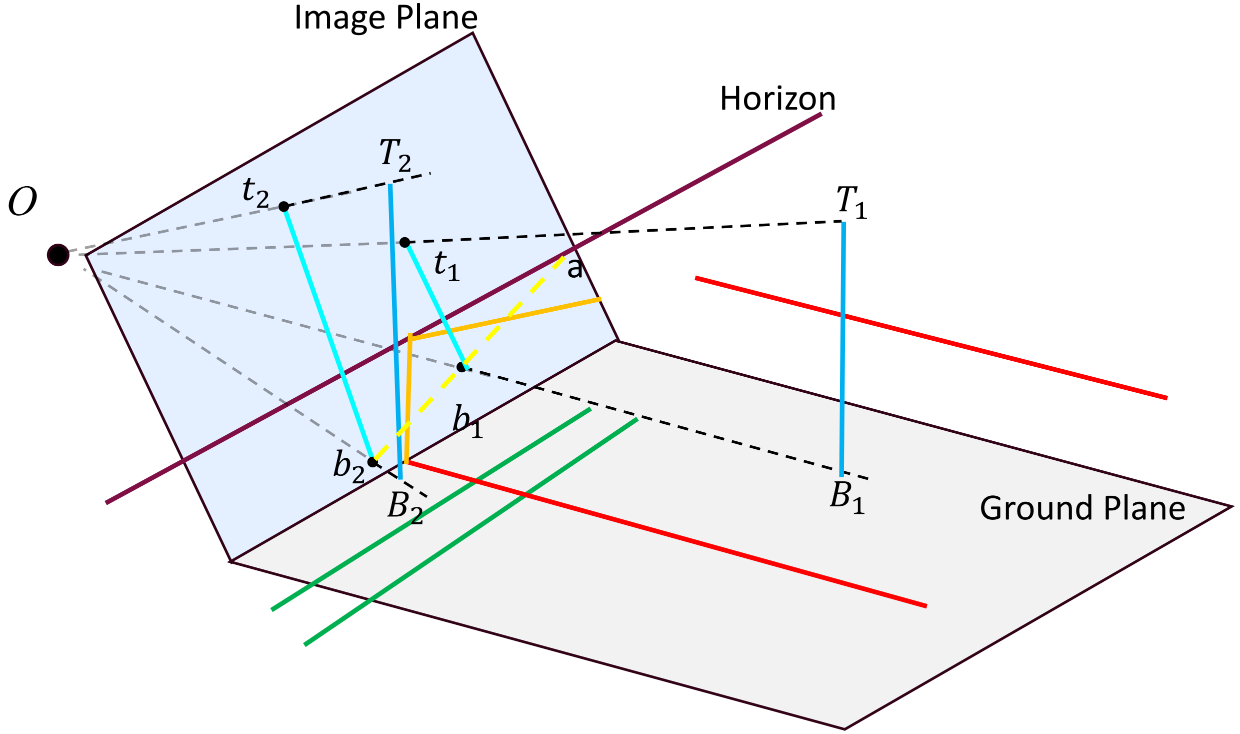
1. Find horizon (from homography, or by finding VPs)



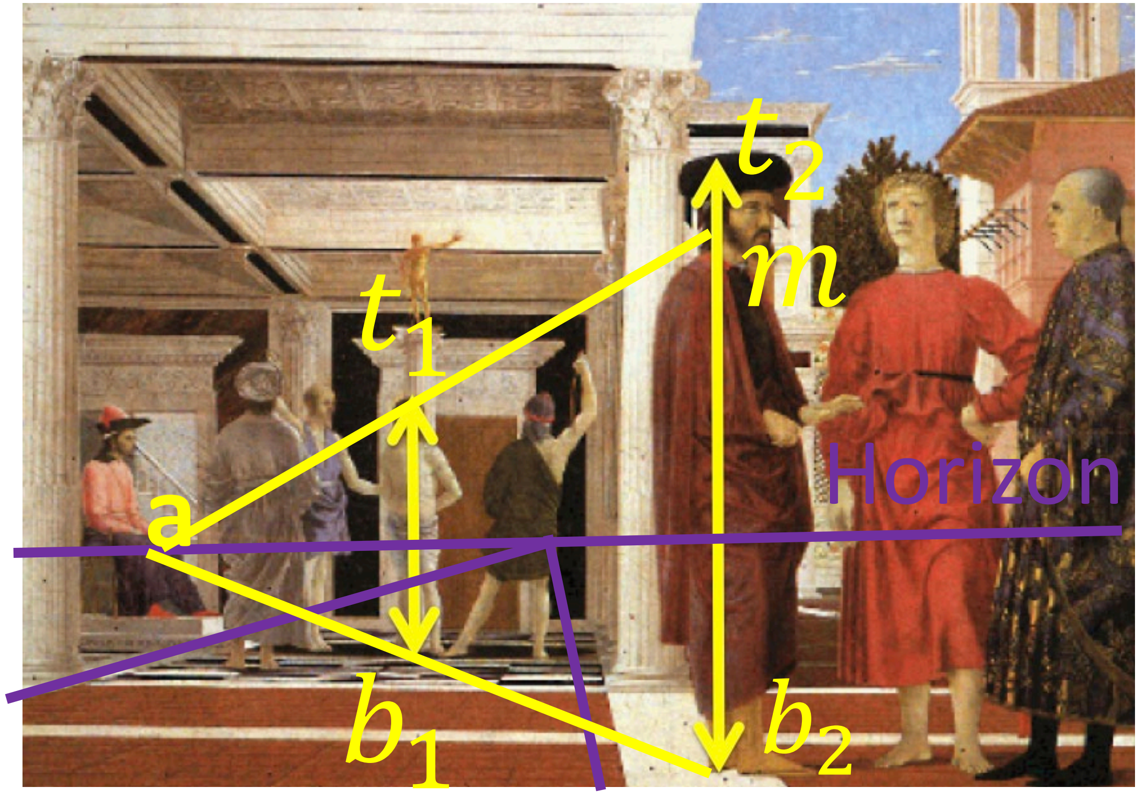


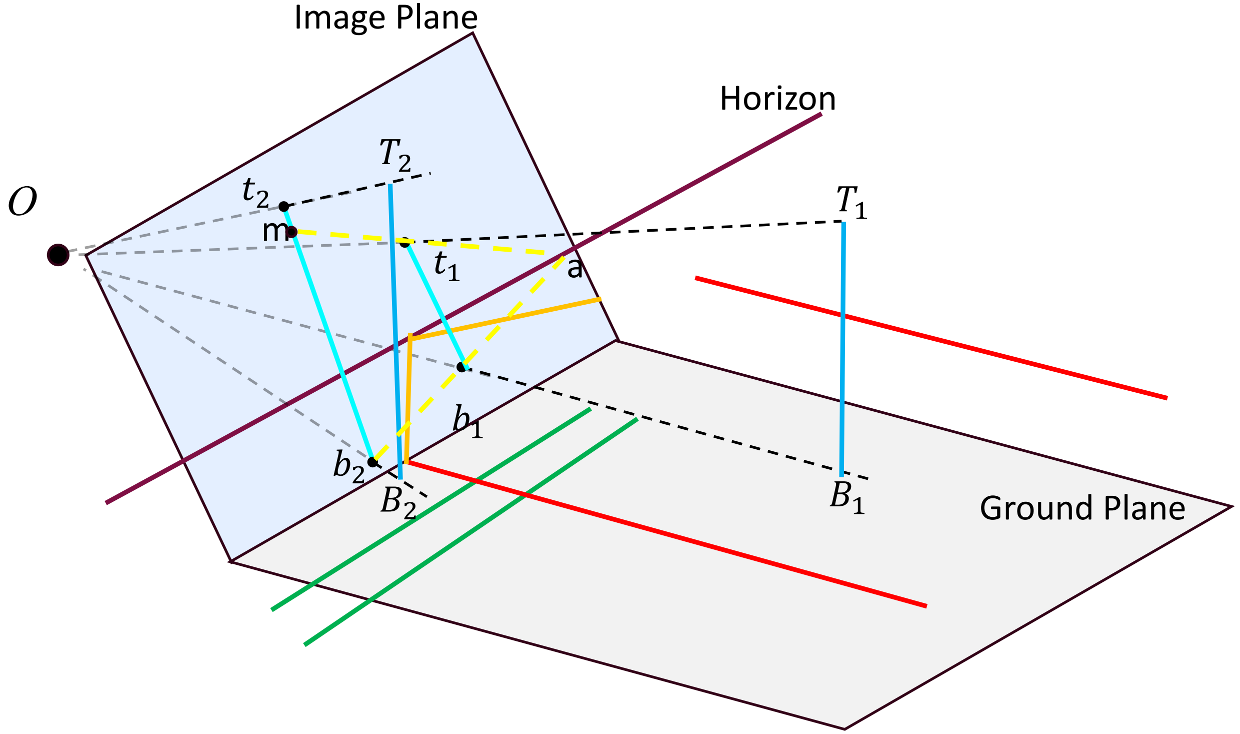
2. Connect the feet of the man and the statue and find intersection a' with horizon!





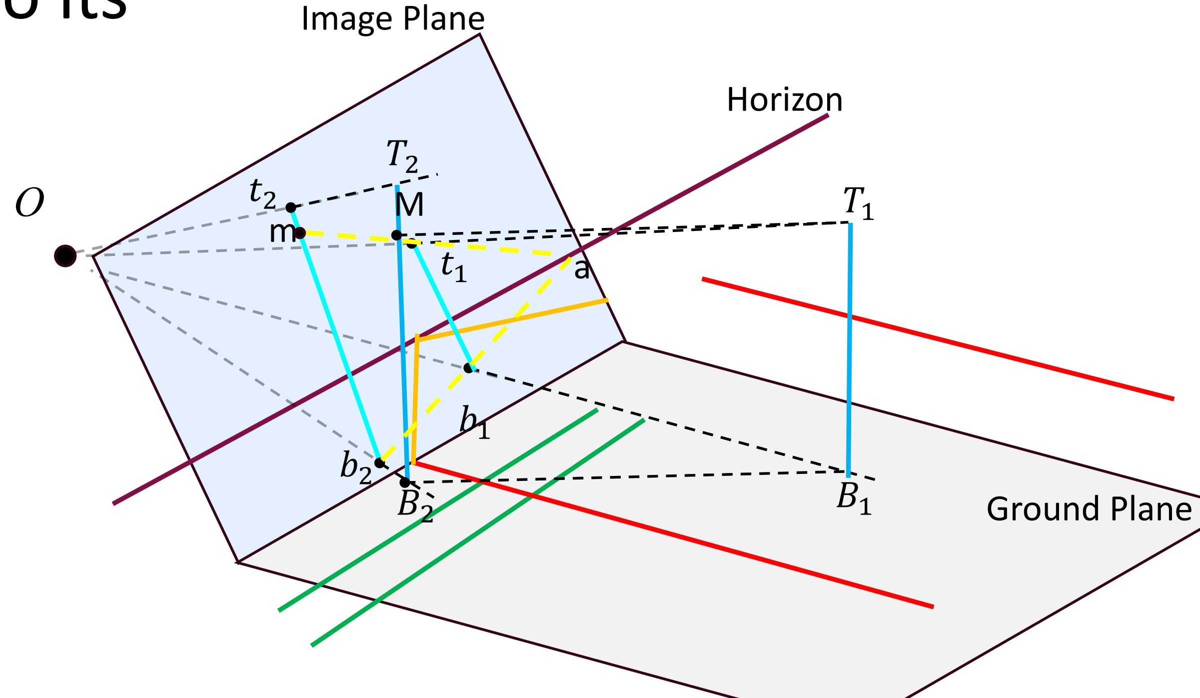
3. Connect a with top of statue t_1



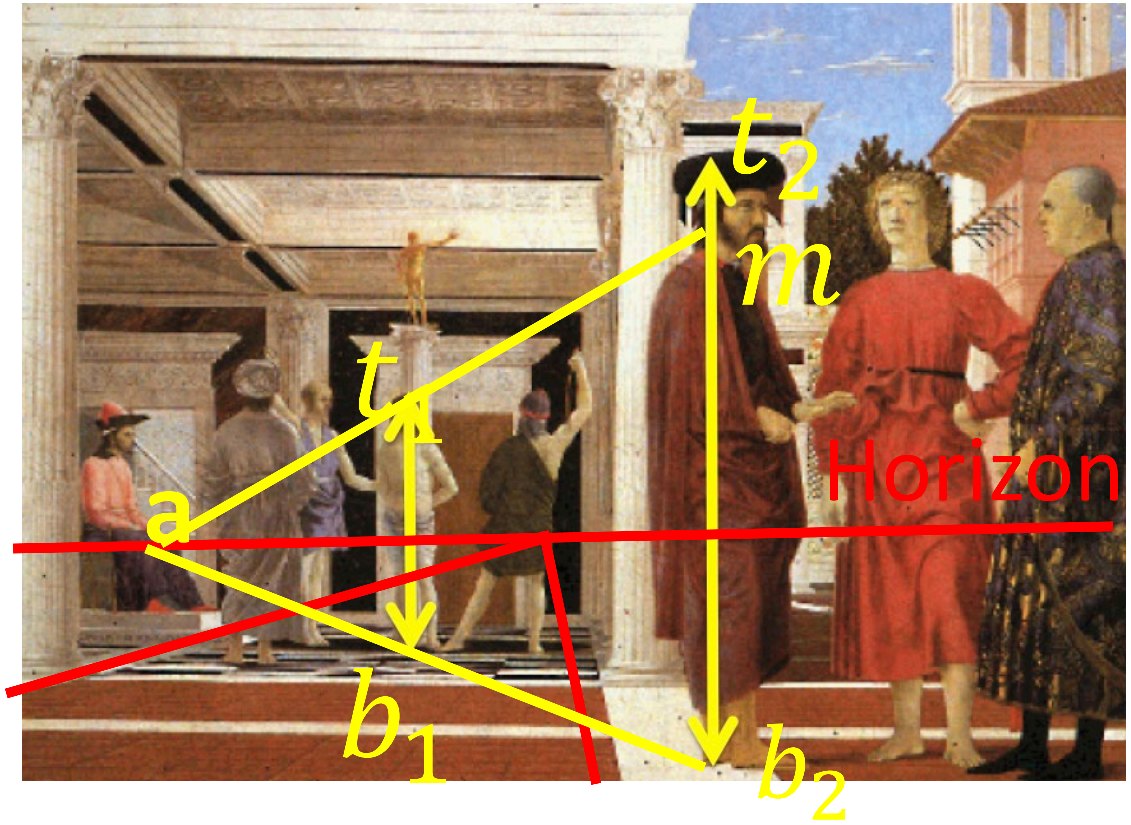


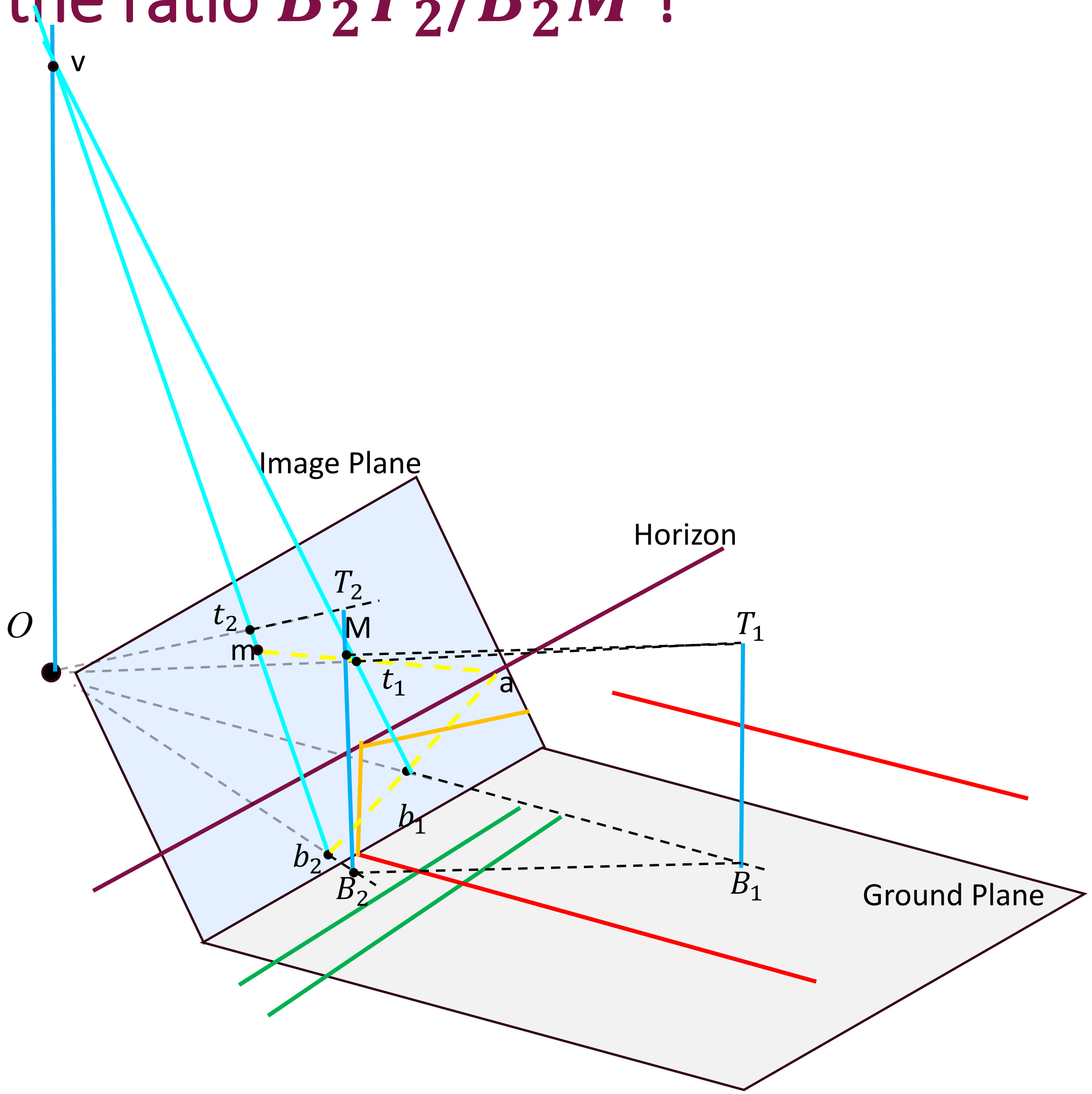
$\boldsymbol{MT_1}$ is parallel to the ground → $B_2M = B_1T_1$

- Consider a world line parallel to B_2B_1 and passing through top of statue T_1. It intersects the person at a point M that is at the same height as T_1. (for practical purposes, we will treat the person and the statue as "vertical lines".)
- The world line MT_1 is parallel to B_2B_1, so its image mt_1 must meet b_2b_1 at its VP=a.
- So $B_2M = B_1T_1$

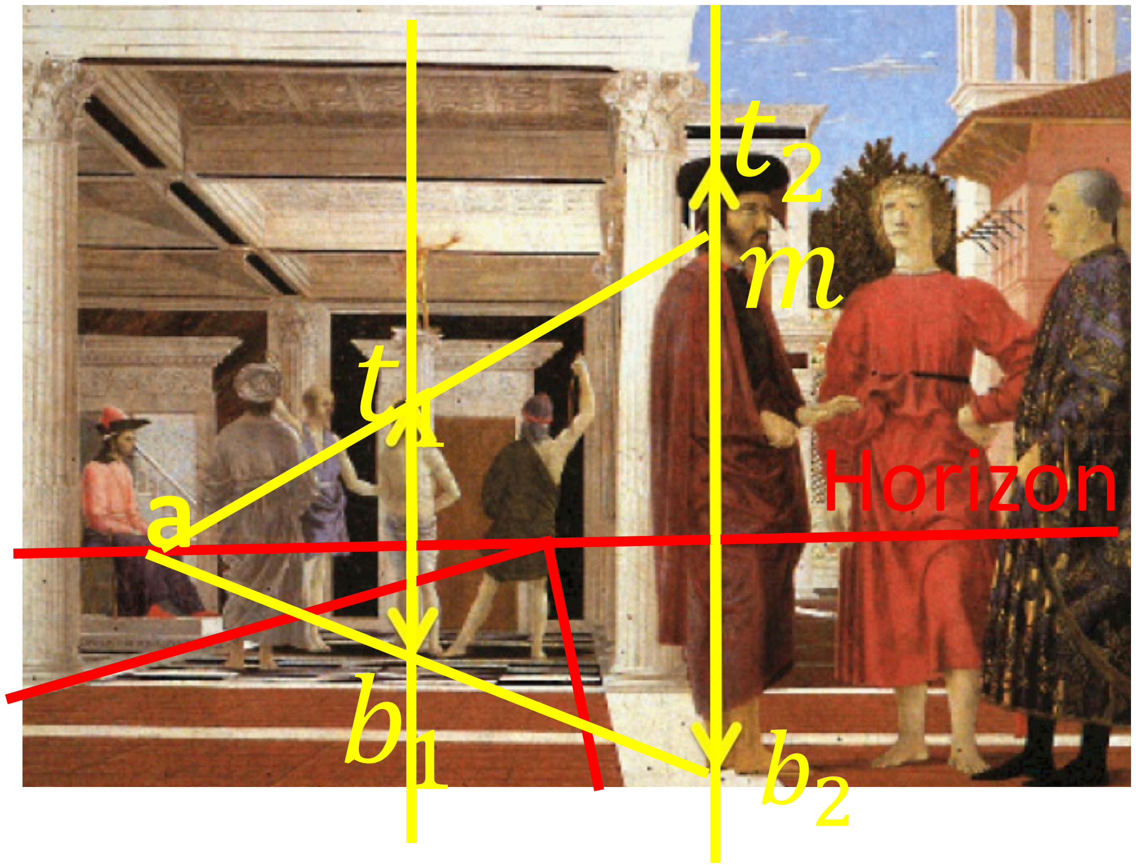


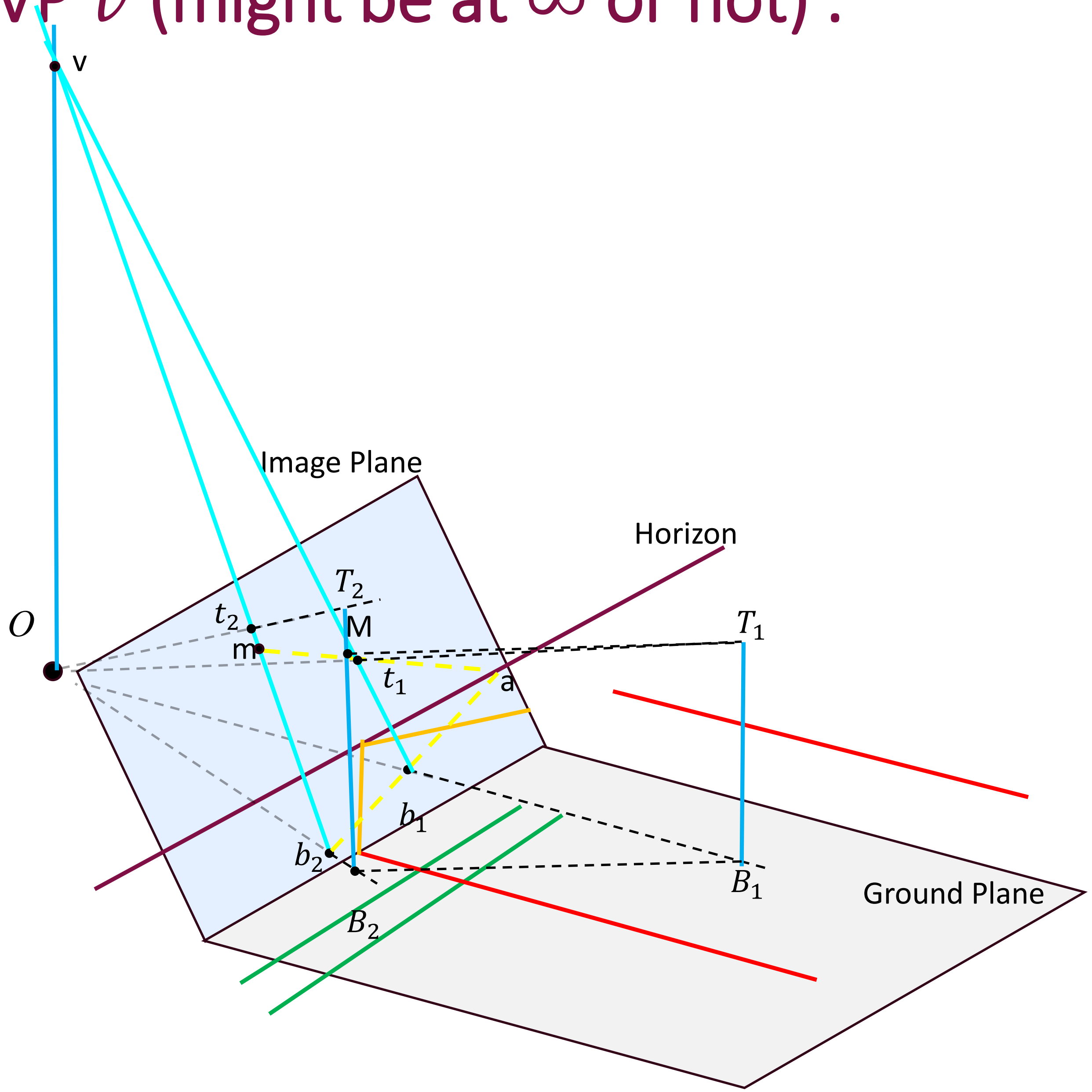
4. But we want B_2T_2 in the world.
How can we compute the ratio B_2T_2/B_2M ?





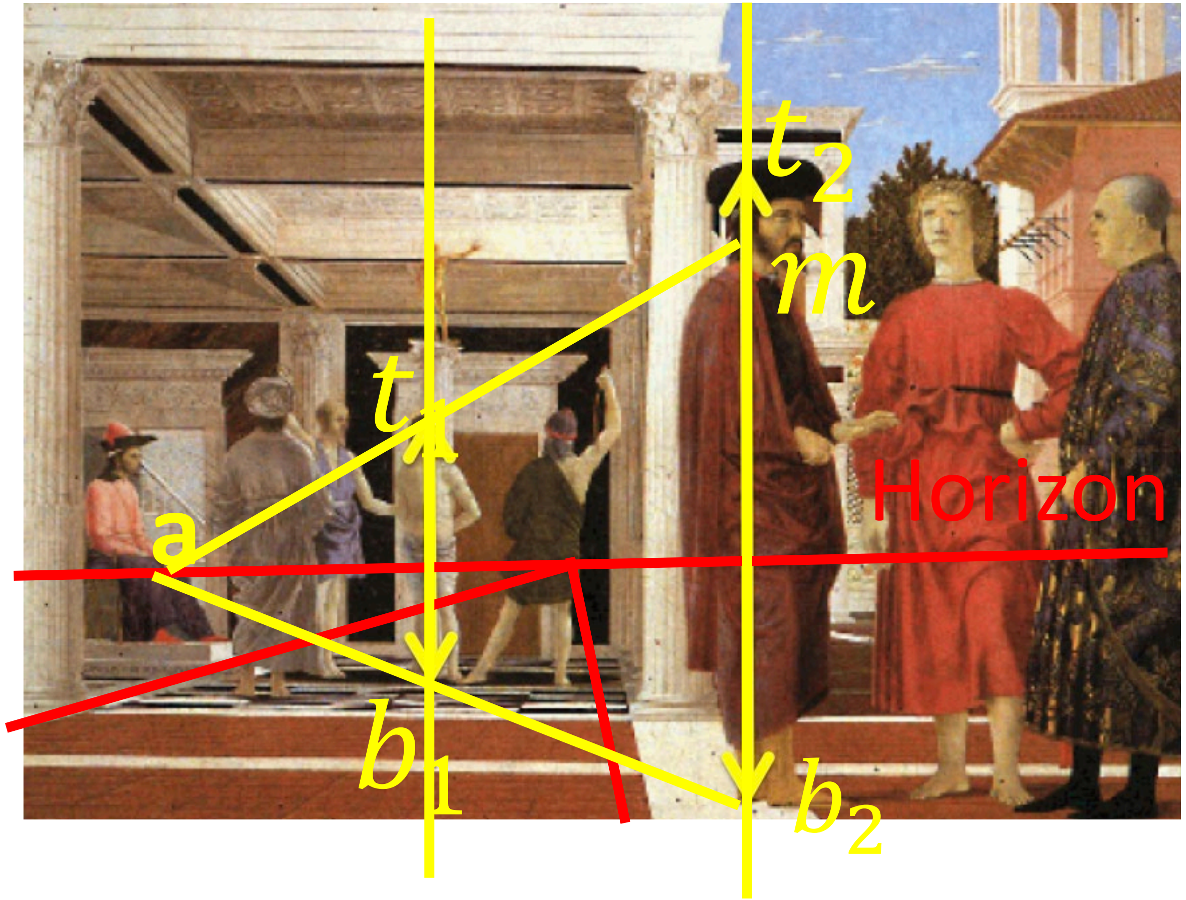
4. Only if we know a vanishing point in the vertical direction. Let $\boldsymbol{b_1 t_1}$ and $\boldsymbol{b_2 m}$ intersect at a vertical VP v (might be at ∞ or not) .

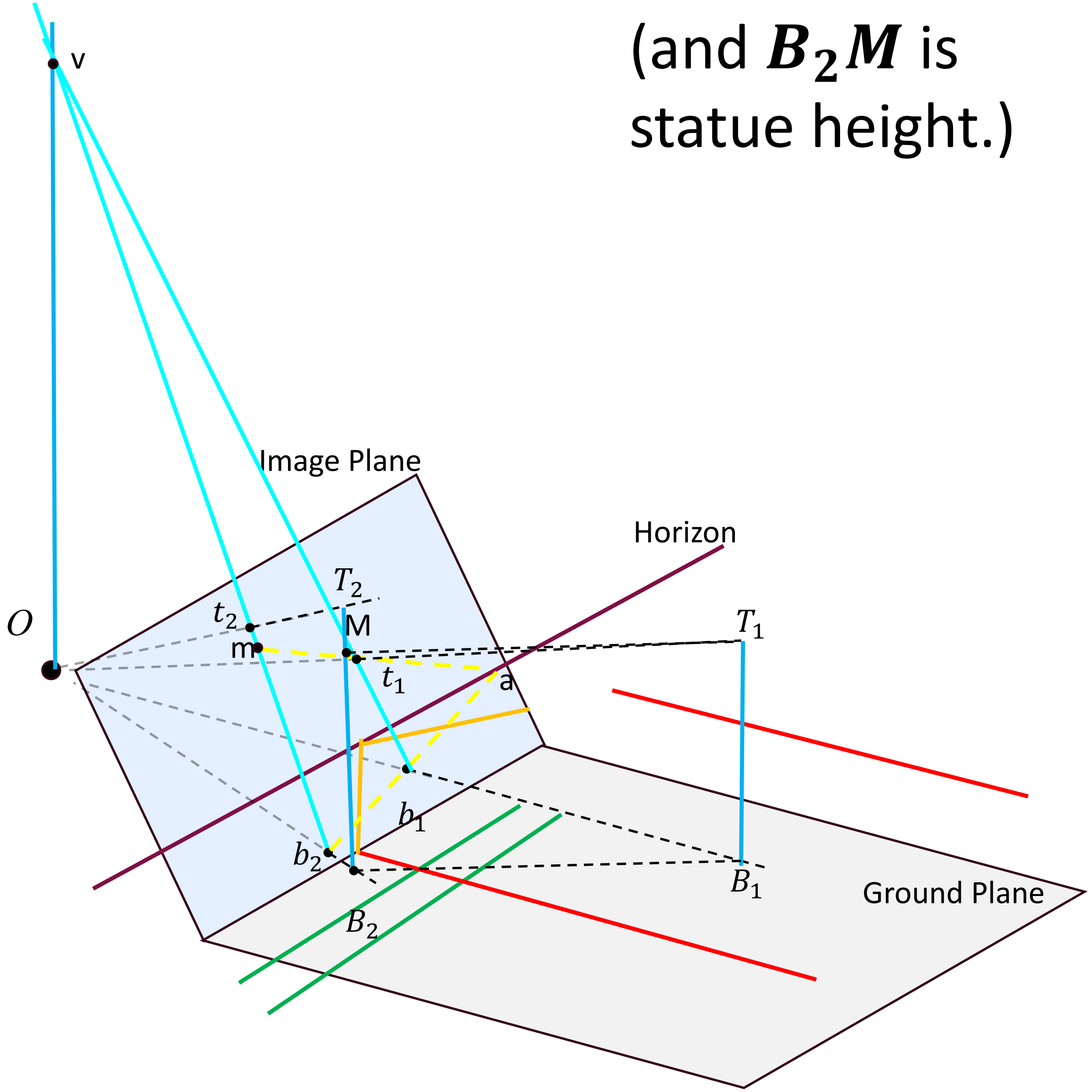




5. Then cross ratio of {v, t2, m, b2} = ratio $B_2T_2 : B_2M$

(and B_2M is statue height.)





Single View Metrology via Cross Ratios

- If we know the vanishing point for a direction, we can compute any ratio along this direction!
- We can transfer lengths among parallel line segments in the world using knowledge of the vanishing point for their direction.
- All of this without explicitly computing any focal length, intrinsics, homographies etc.!

We can do “image forensics” on paintings or old photos!

See also: ZH Sec 8.7,
ZH example 8.25

How to Detect Faked Photos

Techniques that analyze the consistency of elements within an image can help to determine whether it is real or manipulated.

Hany Farid

Fixing Camera 6DoF Pose Estimation w.r.t world plane

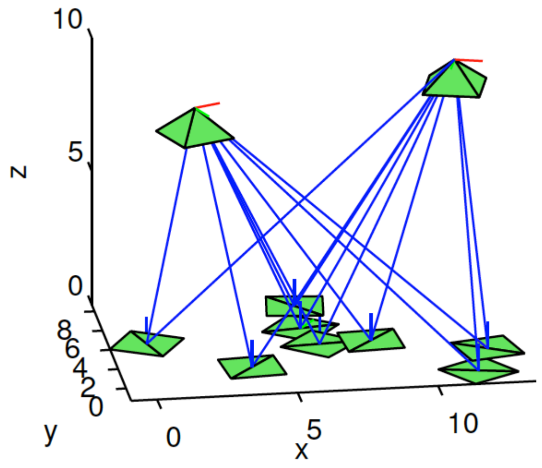
(given homography H & camera intrinsics K)

Camera 6DOF pose

- A camera’s rotation (3DOF) and translation (3DOF) jointly is called its 6-DOF pose.

- “camera pose” estimation = finding the “extrinsics matrix”

Recall: we know how to find homography w.r.t. a planar pattern in the world.



Recall: homography gives pose (given intrinsics K)

Pose from 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 $Z = 0$.

Then the transformation reads

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

And $r_3 = r_1 \times r_2$

But actually, not quite!

- According to the previous slide $K(r_1\ r_2\ T) = H$, or in other words,

$$K^{-1}H = (r_1, r_2, T) \text{ and } r_3 = r_1 \times r_2$$

If only life were so simple!

- Problem: when we ***estimate*** homographies (e.g. through solving linear systems with 2n equations from $n >= 4$ point correspondences), and then compute $K^{-1}H$, we aren't guaranteed to find a *valid* r_1 and r_2 pair. i.e. an orthonormal pair.
    - So, we need to find a way to first "correct" $(K^{-1}H)_{3\times3}$ to get orthonormal r_1 and r_2. Often called the "Procrustes", or "special orthogonal (SO) Procrustes" problem.
    - And we must solve this in real-time for robotics applications, so preferably an inexpensive approach.

The macabre Greek legend of Procrustes

We are trying to get every $(K^{-1}H)$ “traveler” to fit the “bed” of valid rotation matrices by stretching it or chopping it off.

Let us name the columns of $K^{-1}H$:

$$K^{-1}H = \begin{pmatrix} h'_1 & h'_2 & h'_3 \end{pmatrix}$$

We seek orthogonal r_1 and r_2 that are the closest to h'_1 and h'_2. The solution to this problem is given by the Singular Value Decomposition.

We find the orthogonal matrix R that is the closest to $\begin{pmatrix} h'_1 & h'_2 & h'_1 \times h'_2 \end{pmatrix}$:

$$\arg\min_{R \in SO(3)} \| R - \begin{pmatrix} h'_1 & h'_2 & h'_1 \times h'_2 \end{pmatrix} \|_F^2$$

Kabsch algorithm for Procrustes

$$\arg\min_{R \in SO(3)} \| R - \begin{pmatrix} h_1' & h_2' & h_1' \times h_2' \end{pmatrix} \|_F^2$$

If the SVD of

$$\begin{pmatrix} h_1' & h_2' & h_1' \times h_2' \end{pmatrix} = USV^T$$

then the solution is

$$R = U \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \det(UV^T) \end{pmatrix} V^T$$

The diagonal matrix is inserted to guarantee that $\det(R) = 1$.

To find the translation : $T = h_3' / \|h_1'\|$

(In case original columns were not even unit norm)

proof in supp readings- Kabsch-Algorithm-RT-from-H-proof.pdf. We will also prove it in the next class.

Proof in class notes, optional

Full Kabsch algorithm for finding pose via homography

1. Find H up to a scale factor from the point coorrespondences

2. Compute $H' = K^{-1}H$. Let H''s columns be $\begin{pmatrix} a & b & c \end{pmatrix}$

3. Minimize

$$\| \begin{pmatrix} a & b & c \end{pmatrix} - \lambda \begin{pmatrix} r_1 & r_2 & T \end{pmatrix} \|_F$$

w.r.t. $\lambda \in \mathbb{R}, r_1, r_2, T \in \mathbb{R}^3$
s.t. $r_1^T r_2 = 0$ and $\|r_1\| = \|r_2\| = 1$

Let

$$(a\ b) = U_{3x2} \begin{pmatrix} s_1 & 0 \\ 0 & s_2 \end{pmatrix} V_{2x2}^T.$$

Then

$$\begin{pmatrix} r_1 & r_2 \end{pmatrix} = U_{3x2} V_{2x2}^T \quad \text{and} \quad \lambda = \frac{s_1 + s_2}{2}$$

Alternative to running Kabsch including the 3rd column c $= h_1' \times h_2'$ as on last slide

4. $T = c/\lambda$ and $R = \begin{pmatrix} r_1 & r_2 & r_1 \times r_2 \end{pmatrix}$. Scale R to have determinant 1 if needed.

So now, camera pose (actually) known w.r.t world plane!