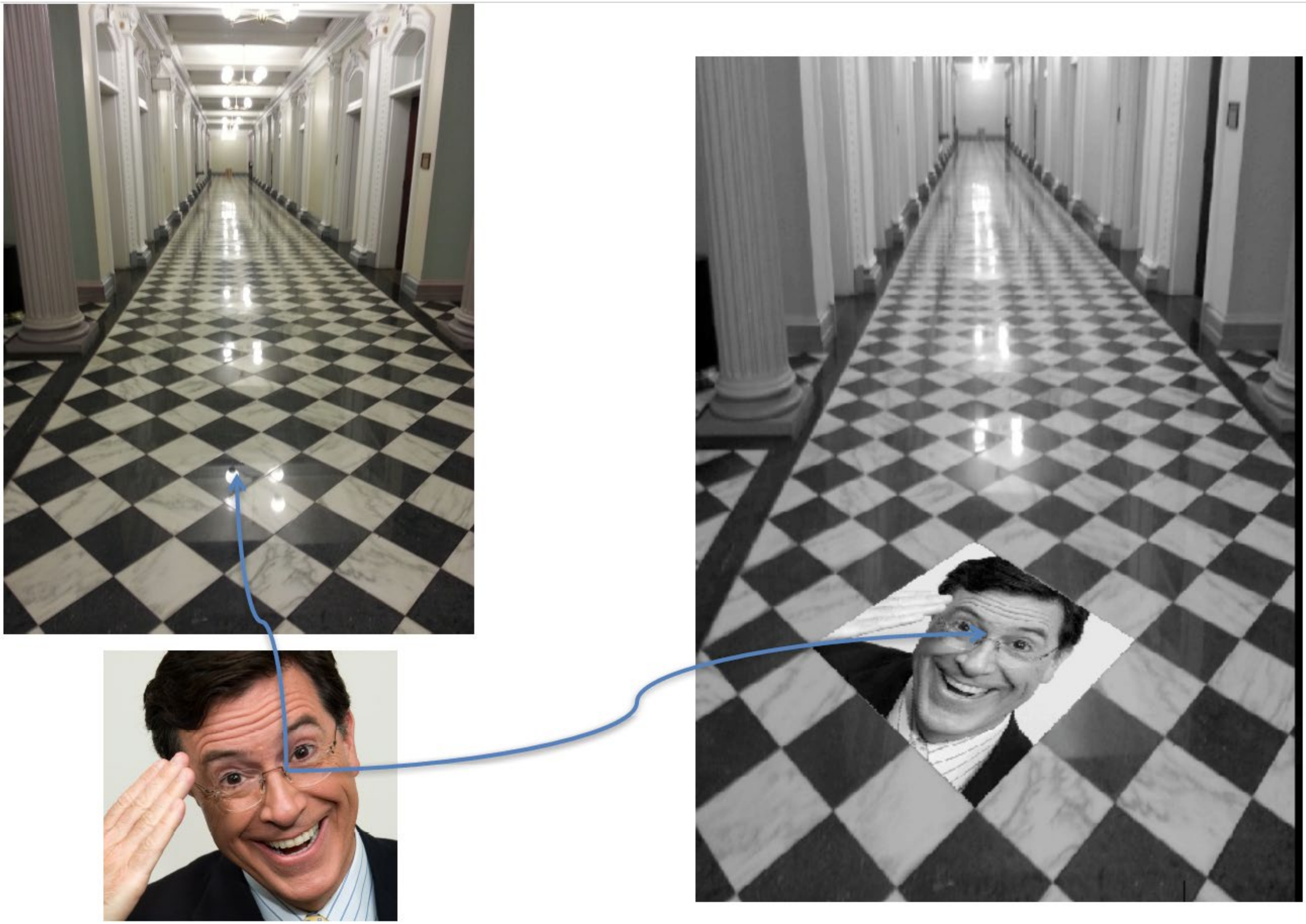
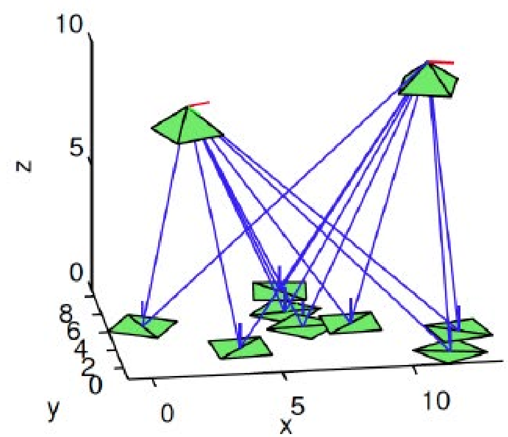
Perception: Pose Estimation

Recall:



Goal:





Pose from Projective Transformations

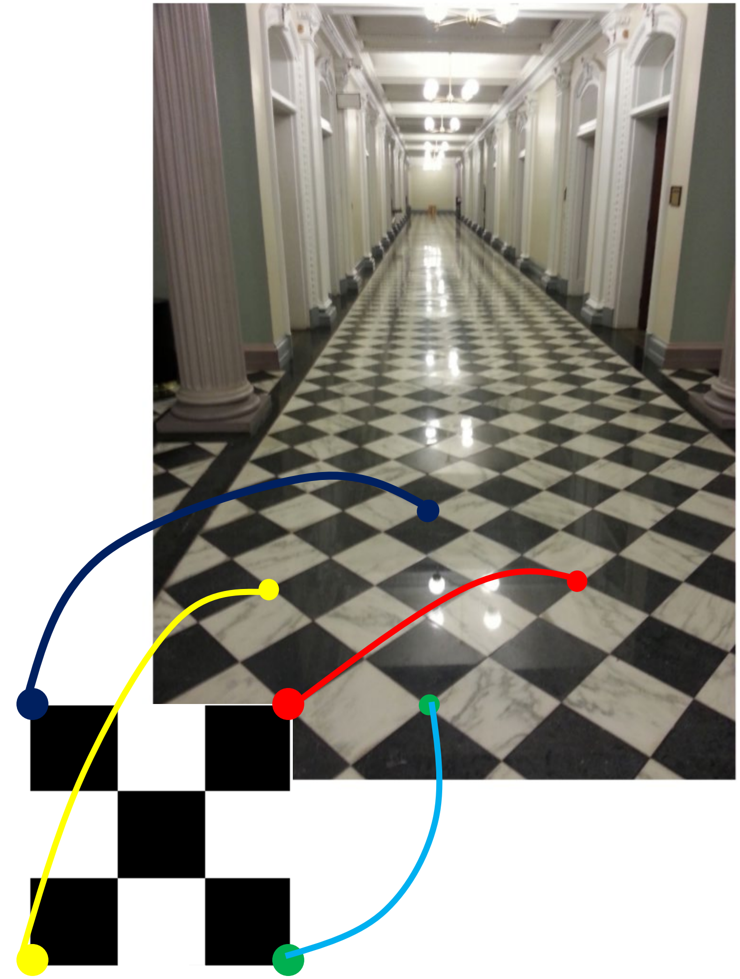
So we have computed the parameters of the projective transformation, representing the transformation of the planar object in the image.

$$\begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix}$$

$$\Big\downarrow ?$$

$$R, T$$

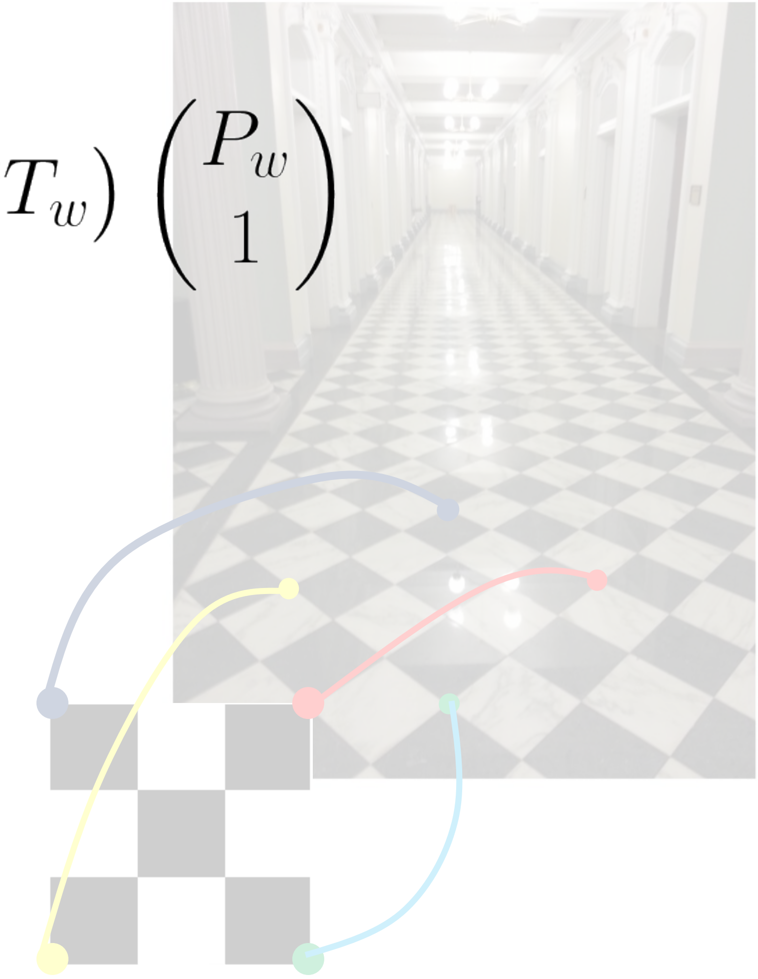
How do we get pose from this?



Derivation

Suppose we are finding our pose relative to points on the ground plane i.e. $Z_w = 0$

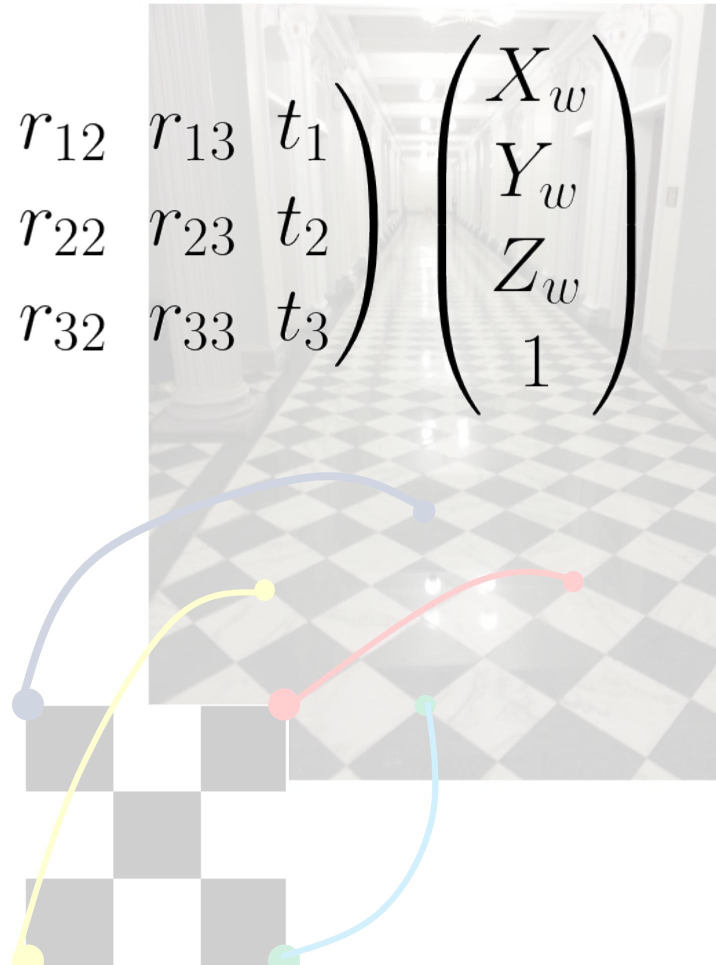
$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \sim K \begin{pmatrix} {}^cR_w & {}^cT_w \end{pmatrix} \begin{pmatrix} P_w \\ 1 \end{pmatrix}$$



Derivation

Suppose we are finding our pose relative to points on the ground plane i.e. $Z_w = 0$

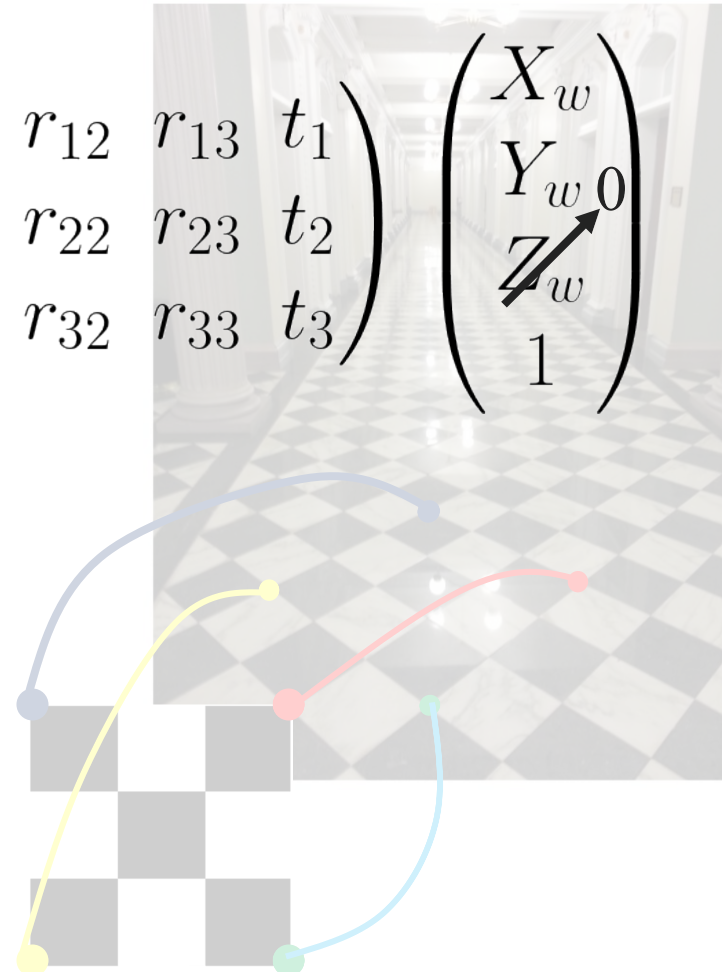
$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \sim \begin{pmatrix} f & 0 & x_0 \\ 0 & f & y_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{pmatrix} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix}$$



Derivation

Suppose we are finding our pose relative to points on the ground plane i.e. $Z_w = 0$

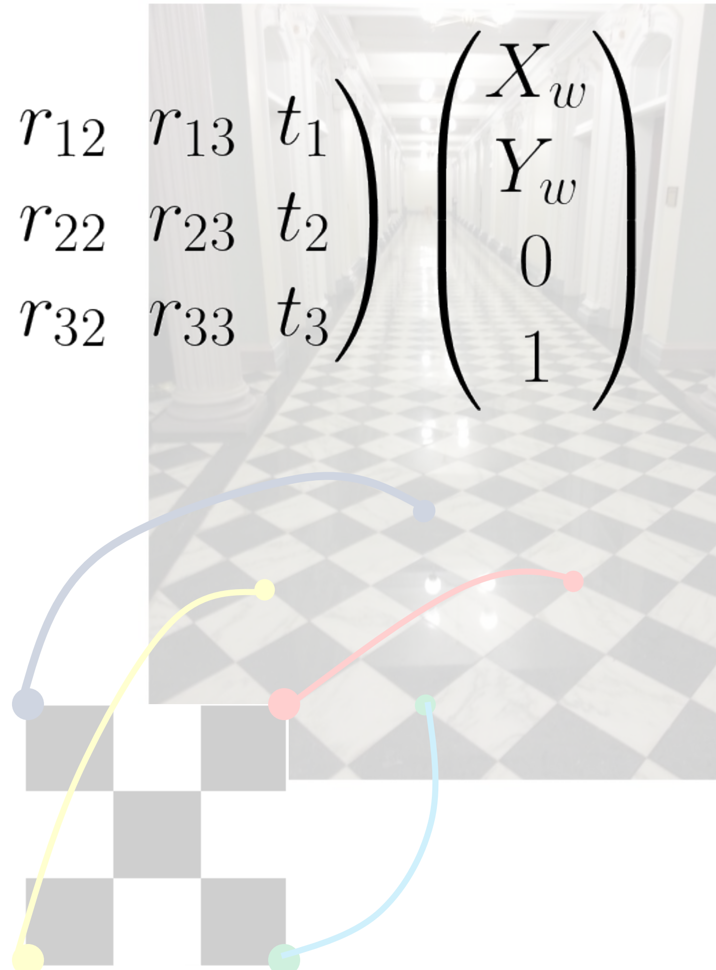
$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \sim \begin{pmatrix} f & 0 & x_0 \\ 0 & f & y_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{pmatrix} \begin{pmatrix} X_w \\ Y_w \\ \cancelto{0}{Z_w} \\ 1 \end{pmatrix}$$



Derivation

Suppose we are finding our pose relative to points on the ground plane i.e. $Z_w = 0$

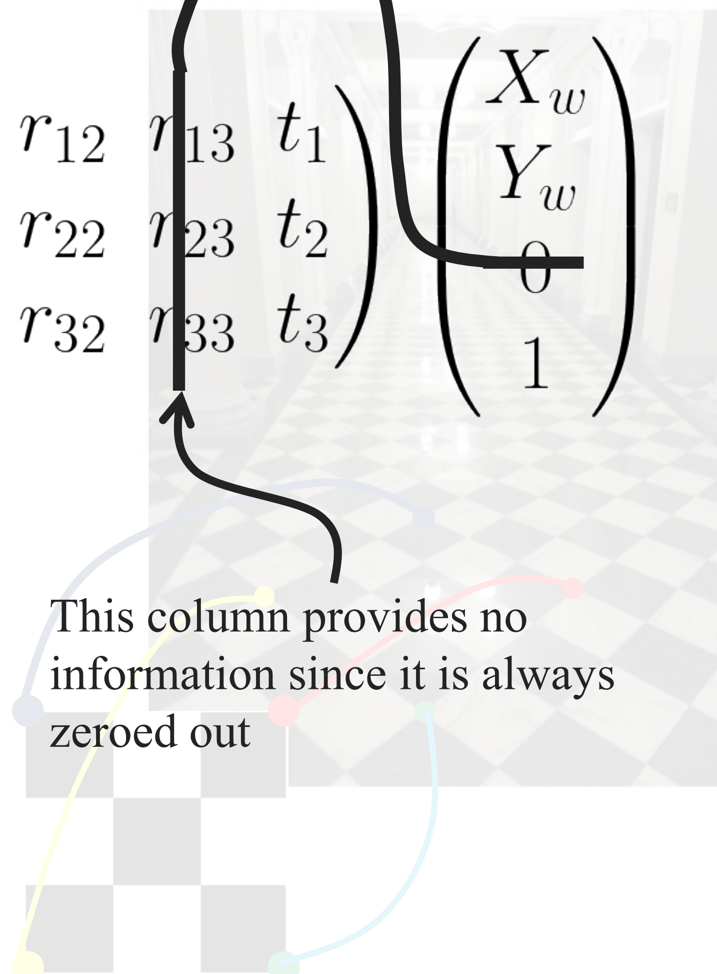
$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \sim \begin{pmatrix} f & 0 & x_0 \\ 0 & f & y_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{pmatrix} \begin{pmatrix} X_w \\ Y_w \\ 0 \\ 1 \end{pmatrix}$$



Derivation

Suppose we are finding our pose relative to points on the ground plane i.e. $Z_w = 0$

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \sim \begin{pmatrix} f & 0 & x_0 \\ 0 & f & y_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{pmatrix} \begin{pmatrix} X_w \\ Y_w \\ 0 \\ 1 \end{pmatrix}$$

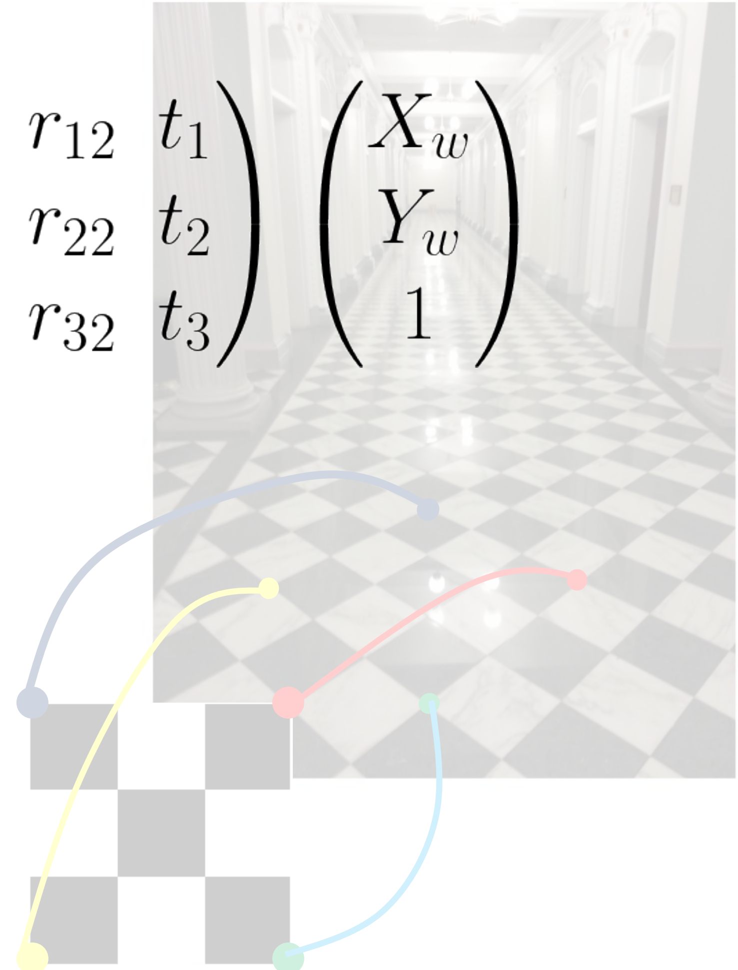


This column provides no information since it is always zeroed out

Derivation

Suppose we are finding our pose relative to points on the ground plane i.e. $Z_w = 0$

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \sim \begin{pmatrix} f & 0 & x_0 \\ 0 & f & y_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & t_1 \\ r_{21} & r_{22} & t_2 \\ r_{31} & r_{32} & t_3 \end{pmatrix} \begin{pmatrix} X_w \\ Y_w \\ 1 \end{pmatrix}$$



Derivation

Suppose we are finding our pose relative to points on the ground plane i.e. $Z_w = 0$

$$\underbrace{\begin{pmatrix} x \\ y \\ 1 \end{pmatrix}}_{\substack{\text{Pixel}\\\text{Coordinates}}} \sim \underbrace{\begin{pmatrix} f & 0 & x_0 \\ 0 & f & y_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & t_1 \\ r_{21} & r_{22} & t_2 \\ r_{31} & r_{32} & t_3 \end{pmatrix}}_{\text{Projective Transformation}} \underbrace{\begin{pmatrix} X_w \\ Y_w \\ 1 \end{pmatrix}}_{\substack{\text{Planar}\\\text{Coordinates}}}$$

$$\begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix}$$

Derivation

Suppose we are finding our pose relative to points on the ground plane i.e. $Z_w = 0$

$$\underbrace{\boxed{\begin{pmatrix} x \\ y \\ 1 \end{pmatrix}}}_{\text{Pixel Coordinates}} \sim \underbrace{\boxed{\begin{pmatrix} f & 0 & x_0 \\ 0 & f & y_0 \\ 0 & 0 & 1 \end{pmatrix}} \begin{pmatrix} r_{11} & r_{12} & t_1 \\ r_{21} & r_{22} & t_2 \\ r_{31} & r_{32} & t_3 \end{pmatrix}}_{\text{Projective Transformation}} \underbrace{\boxed{\begin{pmatrix} X_w \\ Y_w \\ 1 \end{pmatrix}}}_{\text{Planar Coordinates}}$$

$$\boxed{\begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix}}$$

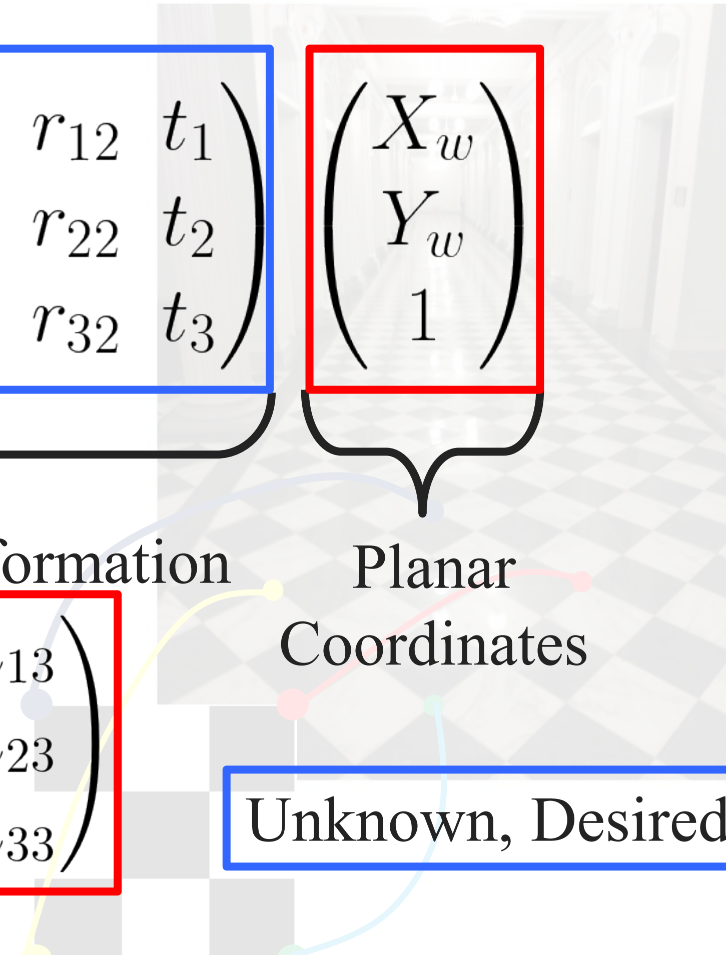
$\boxed{\text{Presumed Known}}$

Derivation

Suppose we are finding our pose relative to points on the ground plane i.e. $Z_w = 0$

$$\underbrace{\boxed{\begin{pmatrix} x \\ y \\ 1 \end{pmatrix}}}_{\text{Pixel Coordinates}} \sim \underbrace{\boxed{\begin{pmatrix} f & 0 & x_0 \\ 0 & f & y_0 \\ 0 & 0 & 1 \end{pmatrix}} \boxed{\begin{pmatrix} r_{11} & r_{12} & t_1 \\ r_{21} & r_{22} & t_2 \\ r_{31} & r_{32} & t_3 \end{pmatrix}}}_{\text{Projective Transformation}} \underbrace{\boxed{\begin{pmatrix} X_w \\ Y_w \\ 1 \end{pmatrix}}}_{\text{Planar Coordinates}}$$

$$\boxed{\begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix}}$$



Presumed Known

Unknown, Desired

Derivation

Now we need to extract the rotation and translation from the remainder

$$\begin{pmatrix} \hat{R}_1 & \hat{R}_2 & \hat{T} \end{pmatrix} = \begin{pmatrix} \hat{r}_{11} & \hat{r}_{12} & \hat{t}_1 \\ \hat{r}_{21} & \hat{r}_{22} & \hat{t}_2 \\ \hat{r}_{31} & \hat{r}_{32} & \hat{t}_3 \end{pmatrix} = \underbrace{\begin{pmatrix} f & 0 & x_0 \\ 0 & f & y_0 \\ 0 & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix}}_{K^{-1}H}$$

Great! We're done right? (nope)

Derivation

Now we need to extract the rotation and translation from the remainder

$$\begin{pmatrix} \hat{R}_1 & \hat{R}_2 & \hat{T} \end{pmatrix} = \begin{pmatrix} \hat{r}_{11} & \hat{r}_{12} & \hat{t}_1 \\ \hat{r}_{21} & \hat{r}_{22} & \hat{t}_2 \\ \hat{r}_{31} & \hat{r}_{32} & \hat{t}_3 \end{pmatrix} = \begin{pmatrix} f & 0 & x_0 \\ 0 & f & y_0 \\ 0 & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix}$$

Need to satisfy constraints:

$$R_1^T R_2 = 0$$
$$\|R_1\| = \|R_2\| = 1$$

Thus we need to modify our estimate to satisfy these constraints

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

Final Solution

Solution for Rotation: Take the Singular Value Decomposition

$$\left(\hat{R}_1 \;\; \hat{R}_2 \;\; \hat{R}_1 \times \hat{R}_2\right) = USV^T$$

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

(Proof of this will be shown later)

To find our estimate of the translation we just make sure it is in the right scale:

$$T = \hat{T}/||\hat{R}_1||$$

Penn Engineering

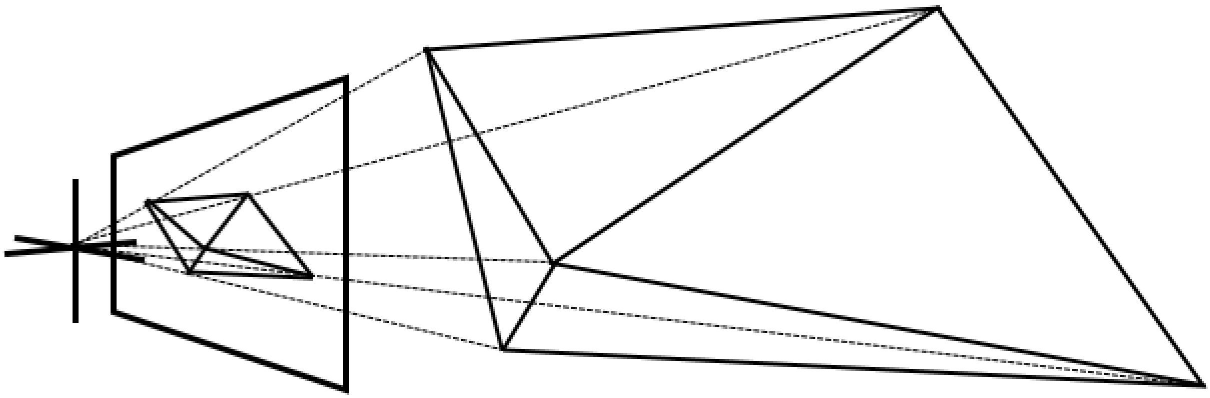
That’s all you need for the project!

Non-Planar Case

We assumed before all the points were on a plane – what if this is not the case? How do we find our pose?

This general problem is known as Perspective N Points or PnP

The smallest N where this can be solved is 3 (why?)



Formal Problem Statement

Given N correspondences of points:

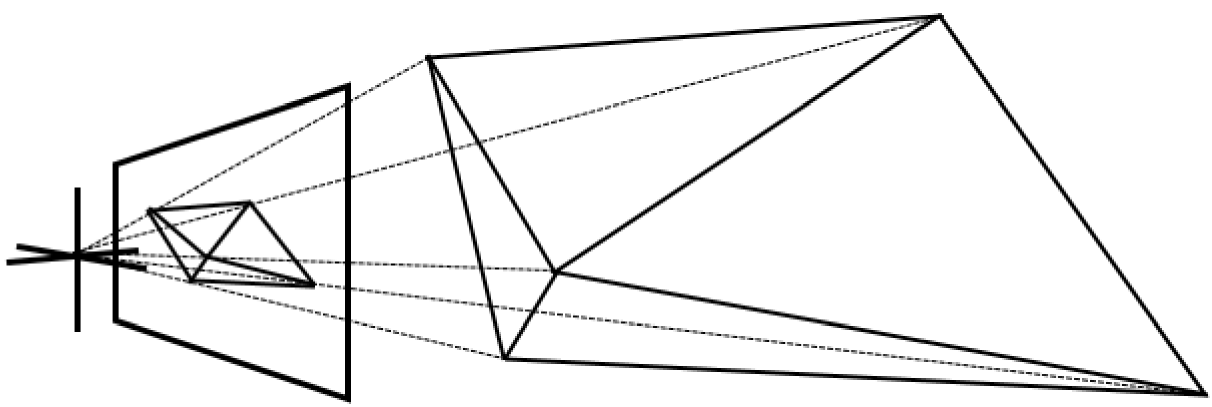
$$(x_i, y_i) \leftrightarrow (X_i, Y_i, Z_i)$$

Find R, T that satisfy (for some λ_i):

$$\lambda_i \begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix} = R \begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix} + T$$

Assumed that 2D coordinates are calibrated, i.e. for pixel coordinates u_i, v_i :

$$\begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix} = K^{-1} \begin{pmatrix} u_i \\ v_i \\ 1 \end{pmatrix}$$



Formal Problem Statement

Given N correspondences of points:

$$(x_i, y_i) \leftrightarrow (X_i, Y_i, Z_i)$$

Find R, T that satisfy (for some λ_i):

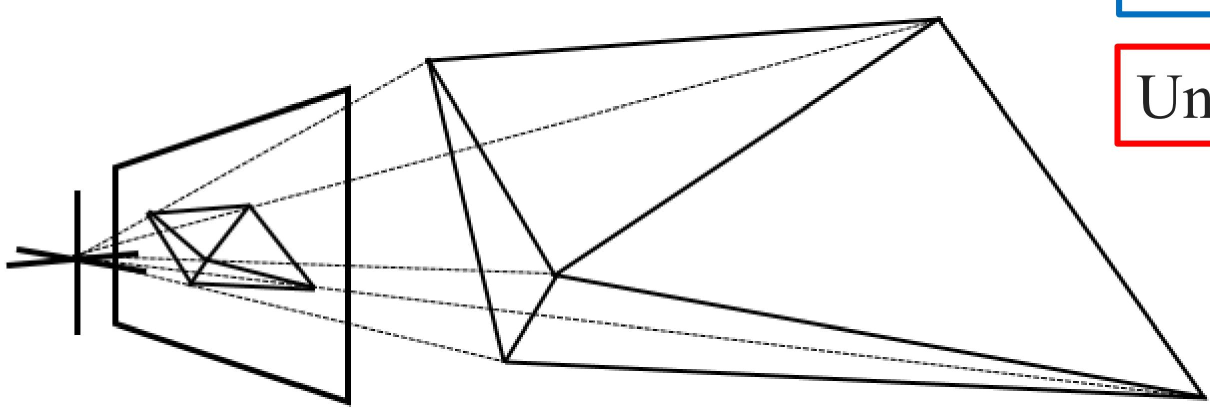
$$\boxed{\lambda_i} \boxed{\begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix}} = \boxed{R} \boxed{\begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix}} + \boxed{T}$$

Assumed that 2D coordinates are calibrated, i.e. for pixel coordinates u_i, v_i :

$$\begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix} = K^{-1} \begin{pmatrix} u_i \\ v_i \\ 1 \end{pmatrix}$$

Known

Unknown



Perspective N Points - PnP

Given N correspondences of points:

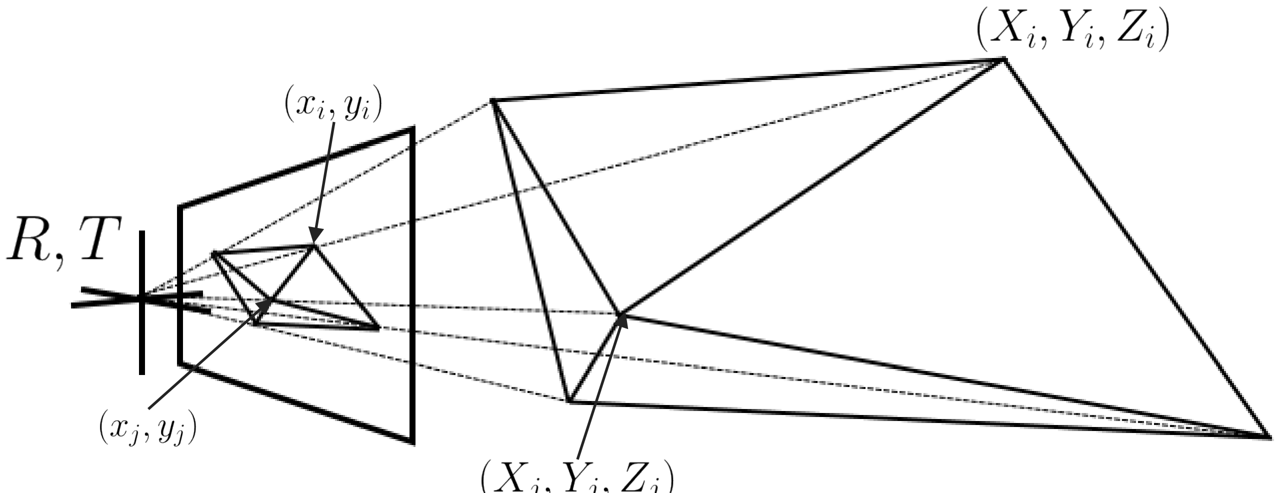
$$(x_i, y_i) \leftrightarrow (X_i, Y_i, Z_i)$$

Find R, T that satisfy (for some λ_i):

$$\lambda_i \begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix} = R \begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix} + T$$

Assumed that 2D coordinates are calibrated, i.e. for pixel coordinates u_i, v_i :

$$\begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix} = K^{-1} \begin{pmatrix} u_i \\ v_i \\ 1 \end{pmatrix}$$

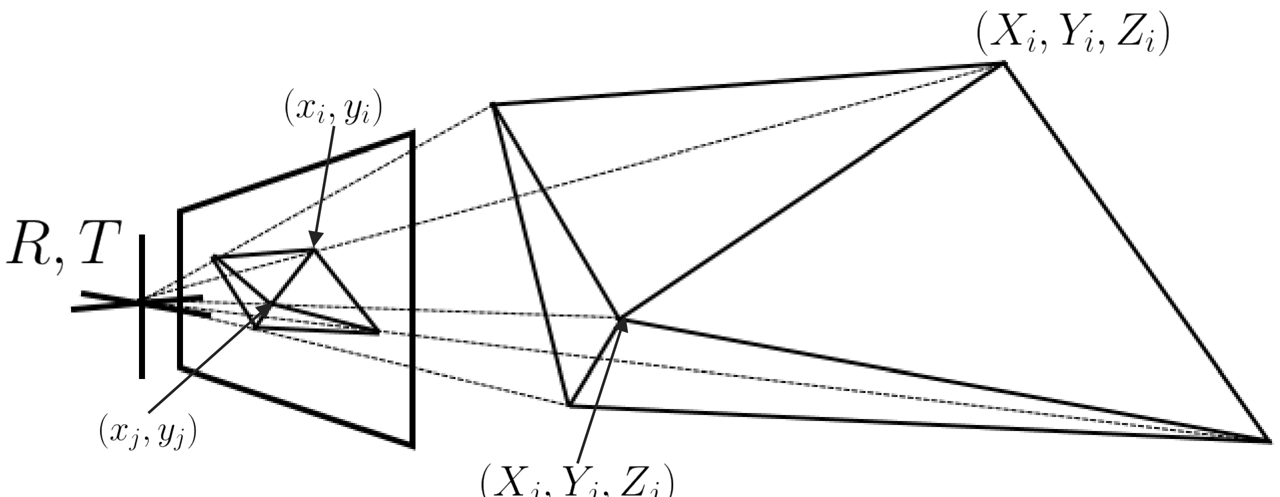


Solution #1: Linear Hack

We could just … ignore the rotation constraint:

$$\arg\min_{\boxed{R\in\mathbb{R}^{3\times 3}}, T\in\mathbb{R}^3, \lambda_i\in\mathbb{R}} \left\| \lambda_i \begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix} - R \begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix} + T \right\|$$

Then use the same trick as before to get the rotation back



Solution #1: Linear Hack

We could just … ignore the rotation constraint:

$$\arg\min_{\boxed{R\in\mathbb{R}^{3\times 3}}, T\in\mathbb{R}^3, \lambda_i\in\mathbb{R}} \left\| \lambda_i \begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix} - R \begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix} + T \right\|$$

Then use the same trick as before to get the rotation back

This is fairly unprincipled approach with not much guaranteeing that the rotation we get is close to the true one … can we do better?

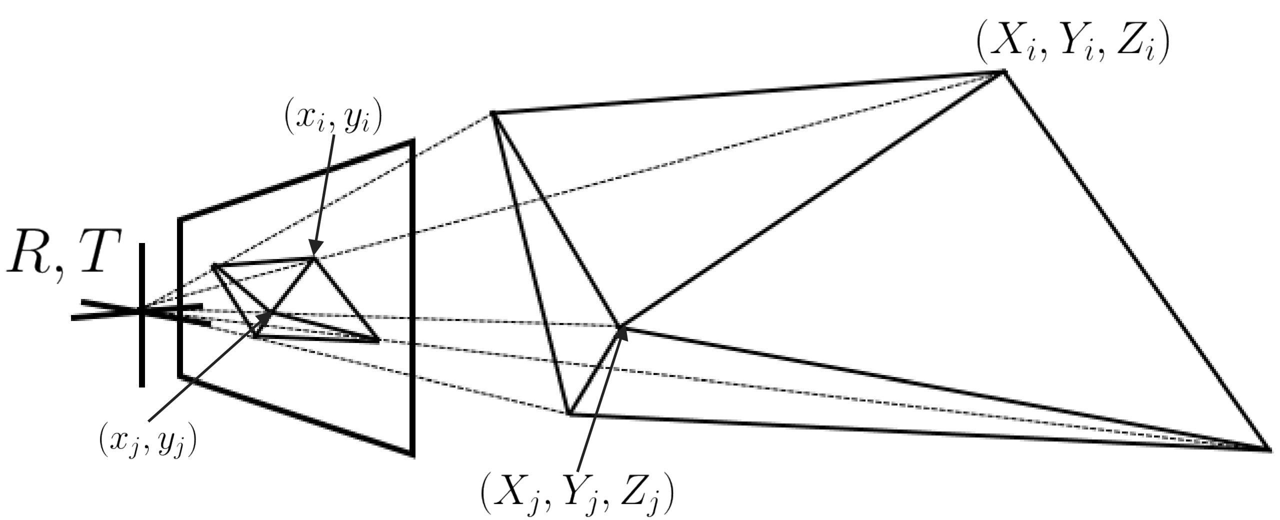
Solution #2: Non-linear Optimization

We could just thrown everything into a large non-linear optimization problem:

$$\arg\min_{R\in SO(3), T\in\mathbb{R}^3, \lambda_i\in\mathbb{R}} \left\| \lambda_i \begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix} - R \begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix} + T \right\|^2$$

Or another way which can get rid of the depths:

$$\arg\min_{R\in SO(3), T\in\mathbb{R}^3} \left\| x_i - \frac{r_{11}X_i + r_{12}Y_i + r_{13}Z_i + T_1}{r_{31}X_i + r_{32}Y_i + r_{33}Z_i + T_3} \right\|^2 + \left\| y_i - \frac{r_{21}X_i + r_{22}Y_i + r_{23}Z_i + T_2}{r_{31}X_i + r_{32}Y_i + r_{33}Z_i + T_3} \right\|^2$$



Penn Engineering

Solution #2: Non-linear Optimization

We could just thrown everything into a large non-linear optimization problem:

$$\arg\min_{\lambda_i \in \mathbb{R},\, R \in SO(3),\, T \in \mathbb{R}^3} \sum_i \left\| \lambda_i \begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix} - R \begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix} + T \right\|^2$$

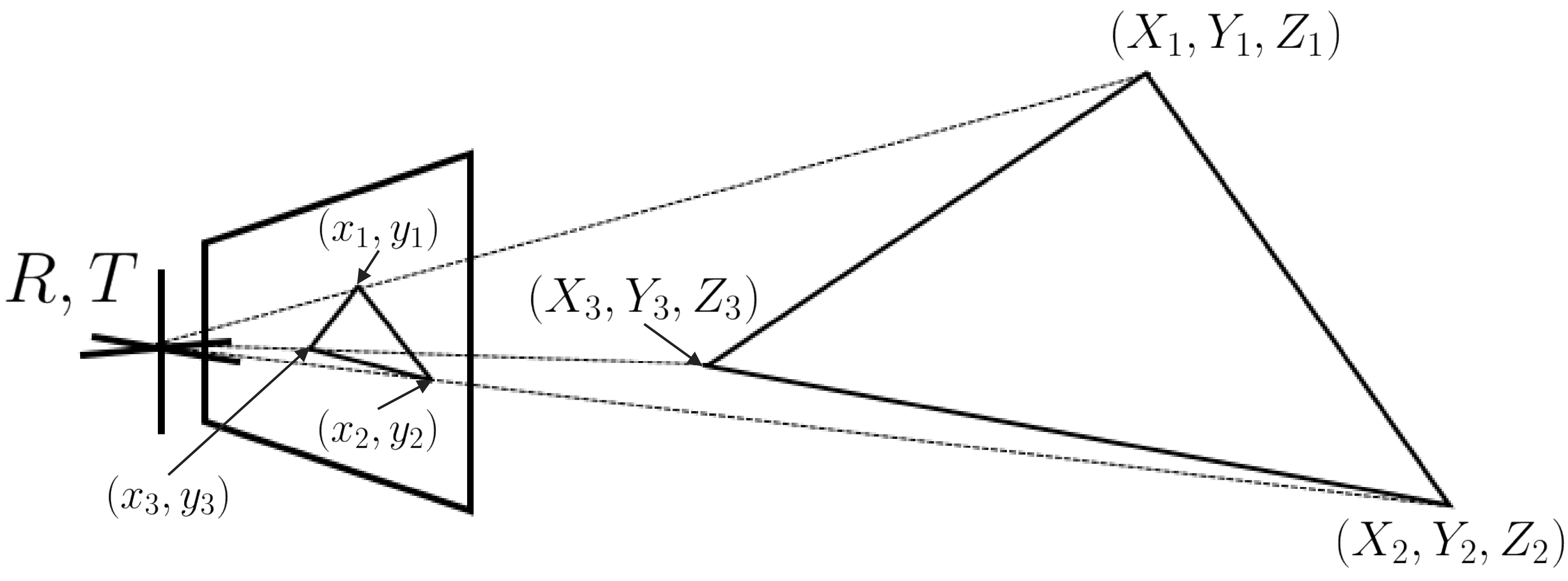
This is a terribly non-convex problem (mainly due to the Rotation constraint) – will need to solve iteratively (e.g. via gradient descent or similar) and not guaranteed to get to global optimum

Or another way which can get rid of the depths:

$$\arg\min_{R \in SO(3),\, T \in \mathbb{R}^3} \sum_i \left\| x_i - \frac{r_{11}X_i + r_{12}Y_i + r_{13}Z_i + T_1}{r_{31}X_i + r_{32}Y_i + r_{33}Z_i + T_3} \right\|^2 + \left\| y_i - \frac{r_{21}X_i + r_{22}Y_i + r_{23}Z_i + T_2}{r_{31}X_i + r_{32}Y_i + r_{33}Z_i + T_3} \right\|^2$$

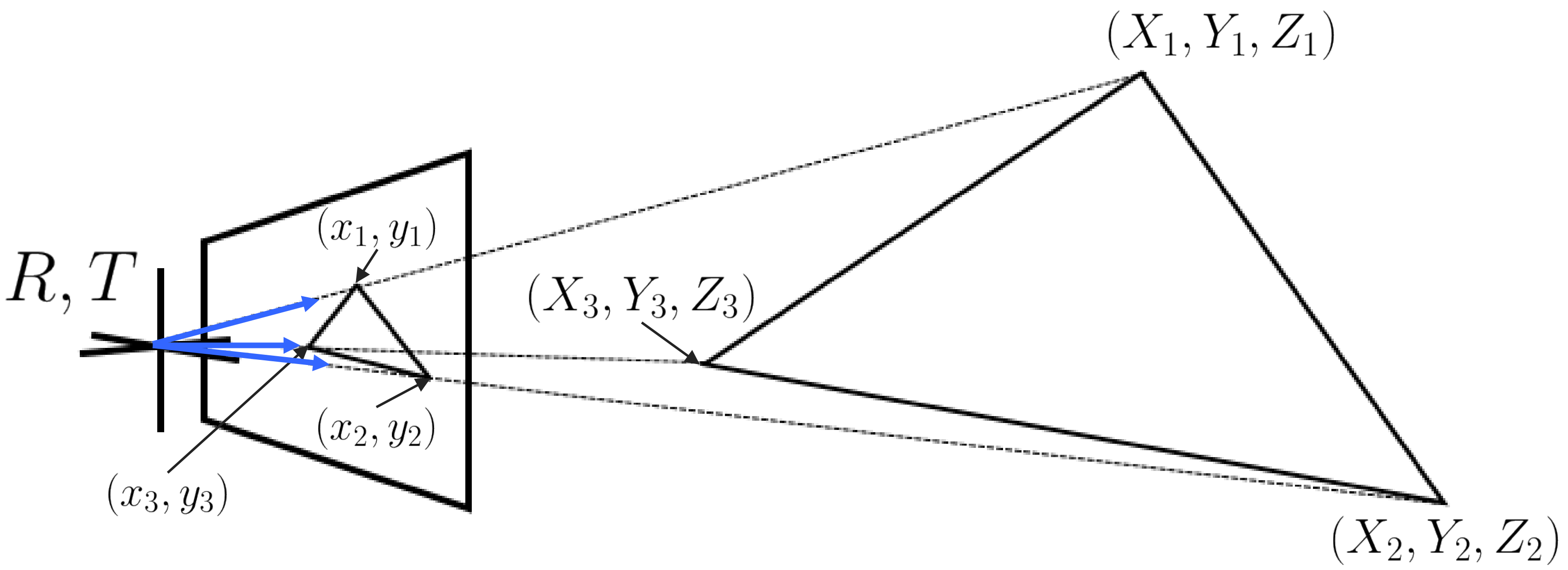
Special Case: P3P

We can try solving for 3 points directly (N=3 case) – if we can solve this quickly, we can take averages or use it as an initialization for the larger non-linear optimization problem. This is a tricky problem, so we will be using a lot of tricks to be able to actually solve it



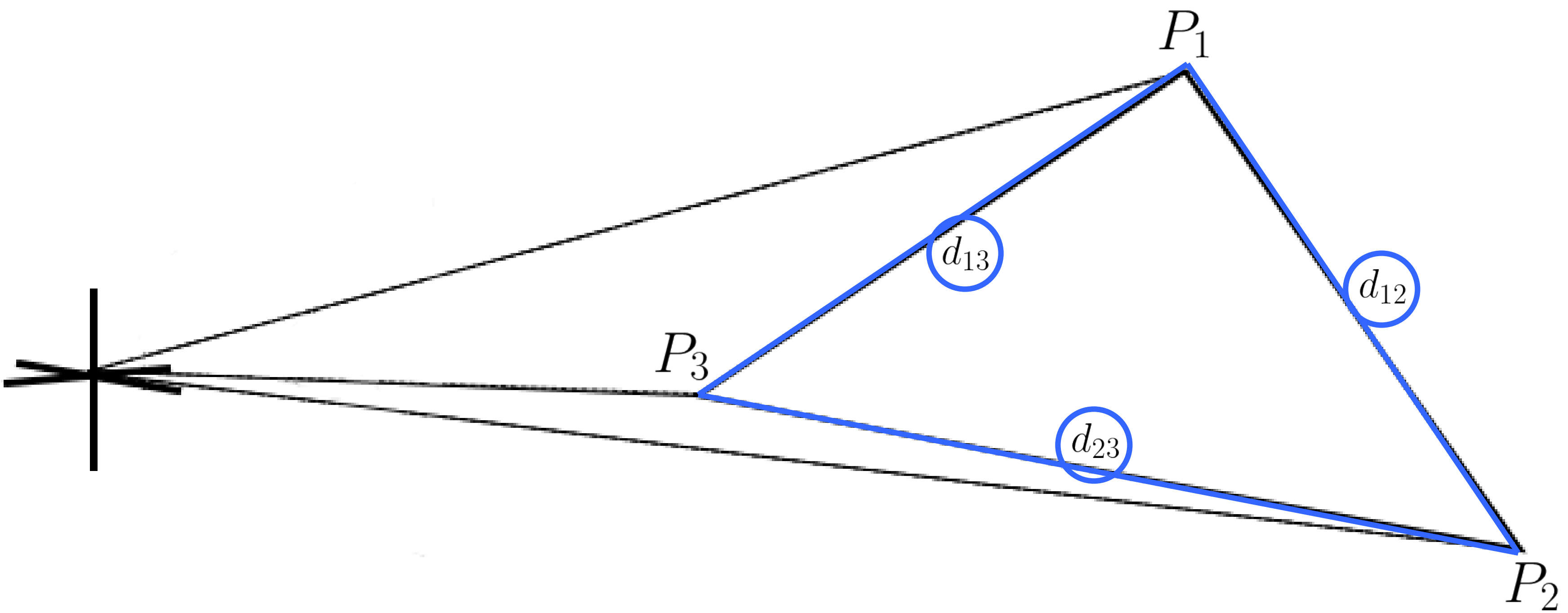
P3P Formulation

First we will use trick here: we use bearings instead of projection points. Then instead of solving for the rotation and translation directly, solve for the distances along these bearings to the 3D points they correspond to



P3P Formulation

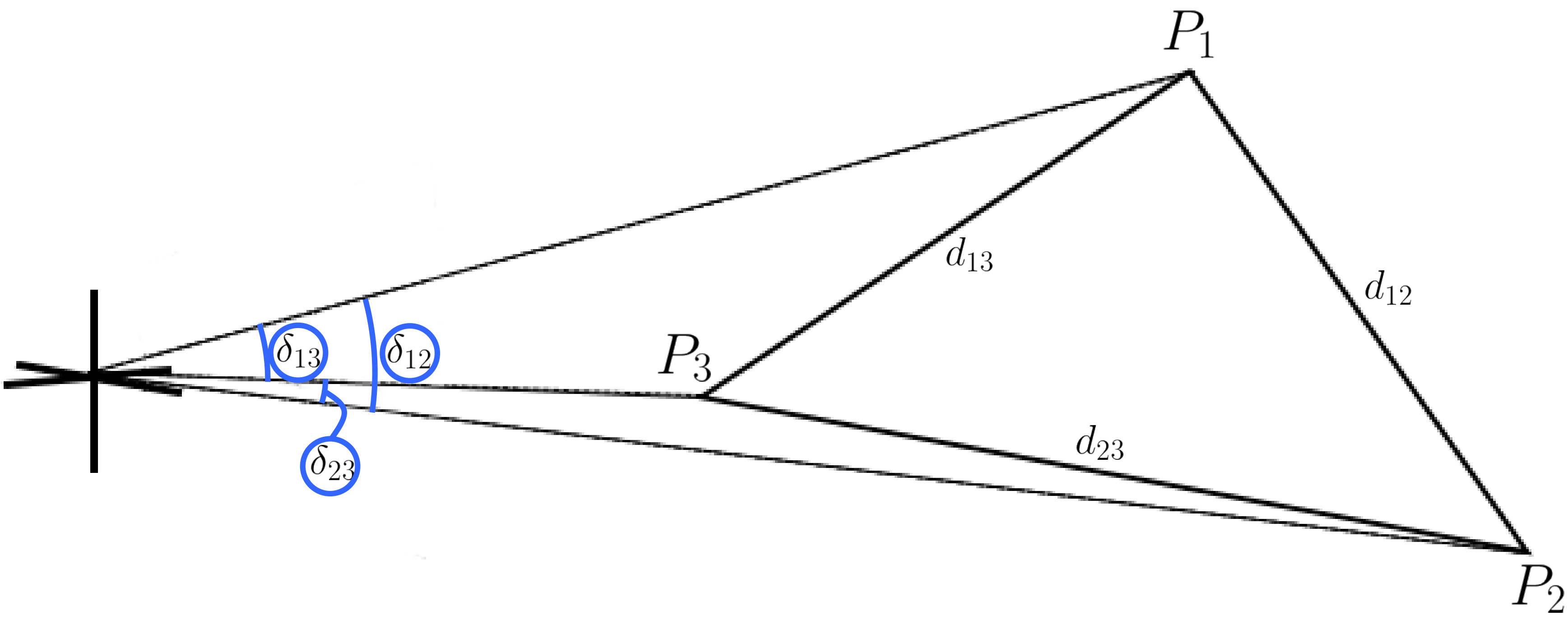
Since we know the 3D position of the points P_i we can calculate the distance between points *i* and *j*, denoted d_{ij}



Penn Engineering

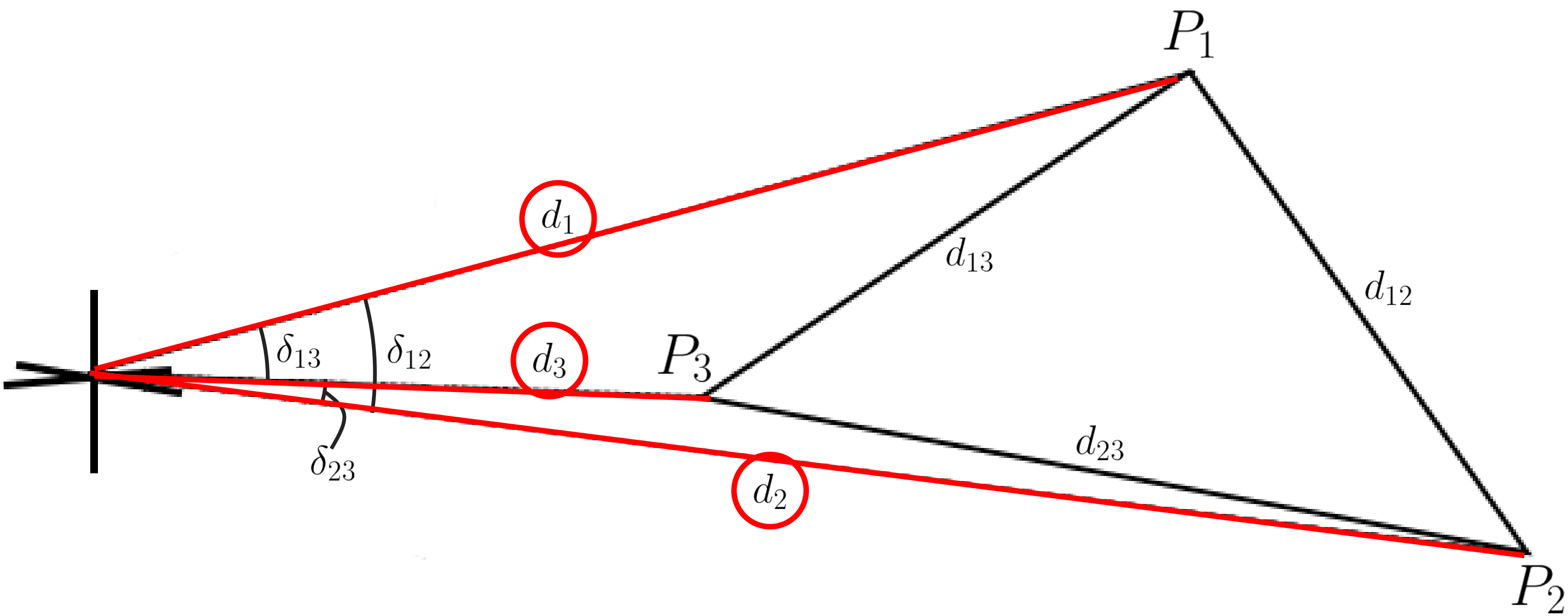
P3P Formulation

We also can calculate the angle between bearings of points i and j, denoted δ_{ij}



P3P Formulation

Call the distances along the bearings i.e. distance from the camera origin to the 3D points. These are the unknowns we are solving for

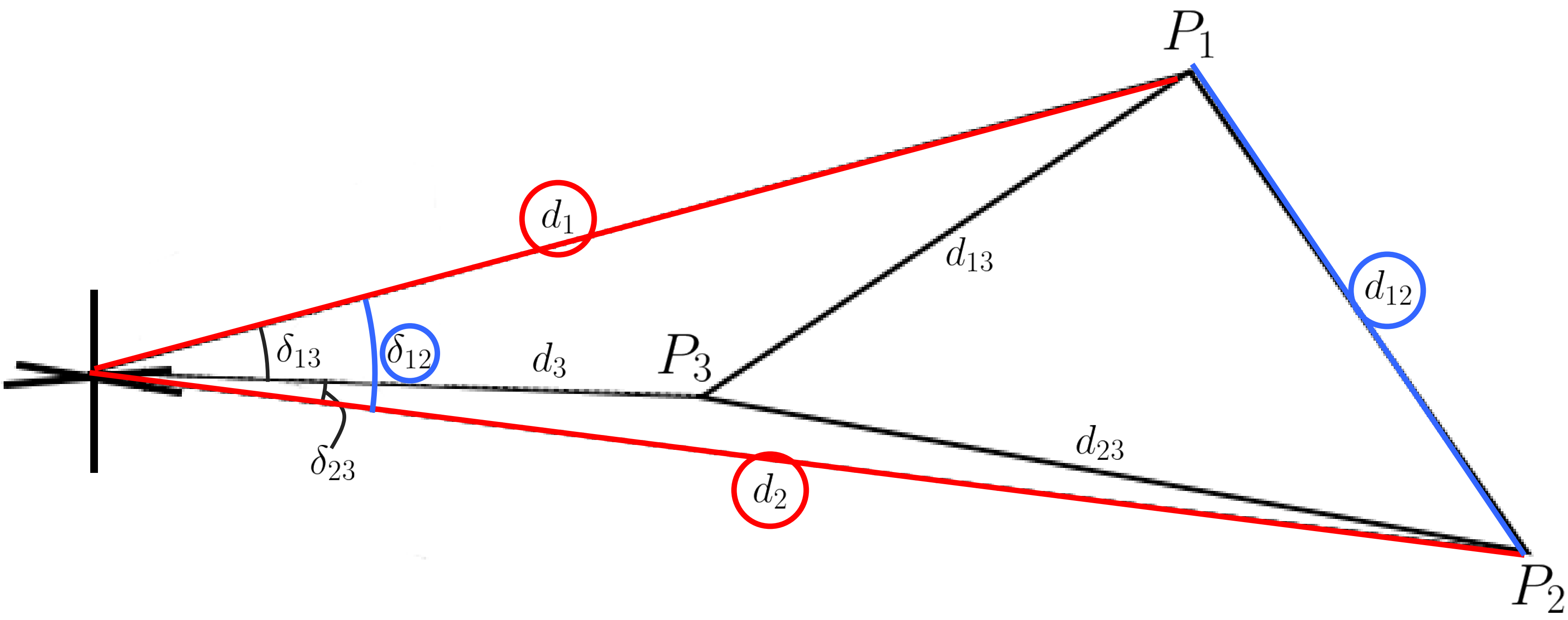


Penn Engineering

Solving P3P

Notice we have three triangles. Pulling out some trigonometric identities we come at the law of cosines (3 of them):

$$d_i^2 + d_j^2 - 2 d_i d_j \cos(\delta_{ij}) = d_{ij}^2$$



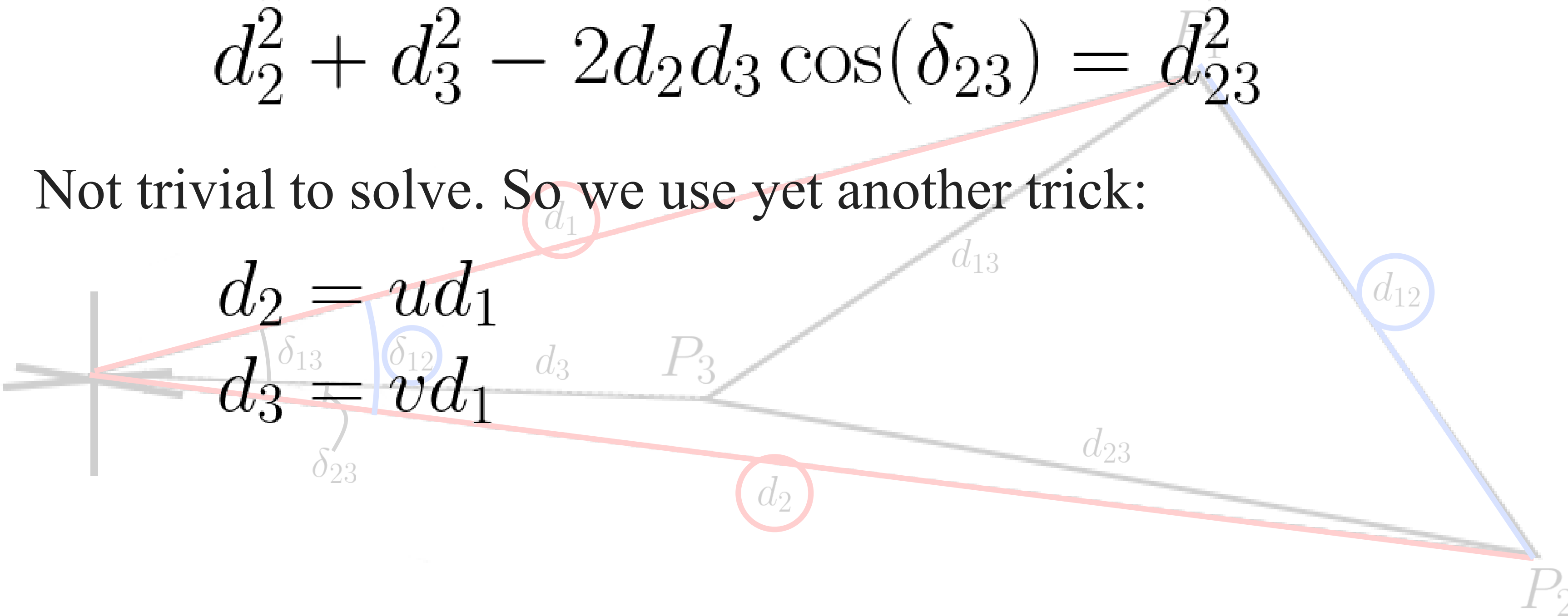
Solving P3P

So now we have to solve for 3 unknowns, and we 3 multivariate quadratic equations:

$$
\begin{aligned}
d_1^2 + d_2^2 - 2d_1 d_2 \cos(\delta_{12}) &= d_{12}^2 \\
d_1^2 + d_3^2 - 2d_1 d_3 \cos(\delta_{13}) &= d_{13}^2 \\
d_2^2 + d_3^2 - 2d_2 d_3 \cos(\delta_{23}) &= d_{23}^2
\end{aligned}
$$

Not trivial to solve. So we use yet another trick:

$$
\begin{aligned}
d_2 &= u d_1 \\
d_3 &= v d_1
\end{aligned}
$$



Solving P3P

So now we have to solve for 3 unknowns, and we 3 multivariate quadratic equations:

$$d_2^2 + d_3^2 - 2d_2d_3\cos(\delta_{23}) = d_{23}^2$$

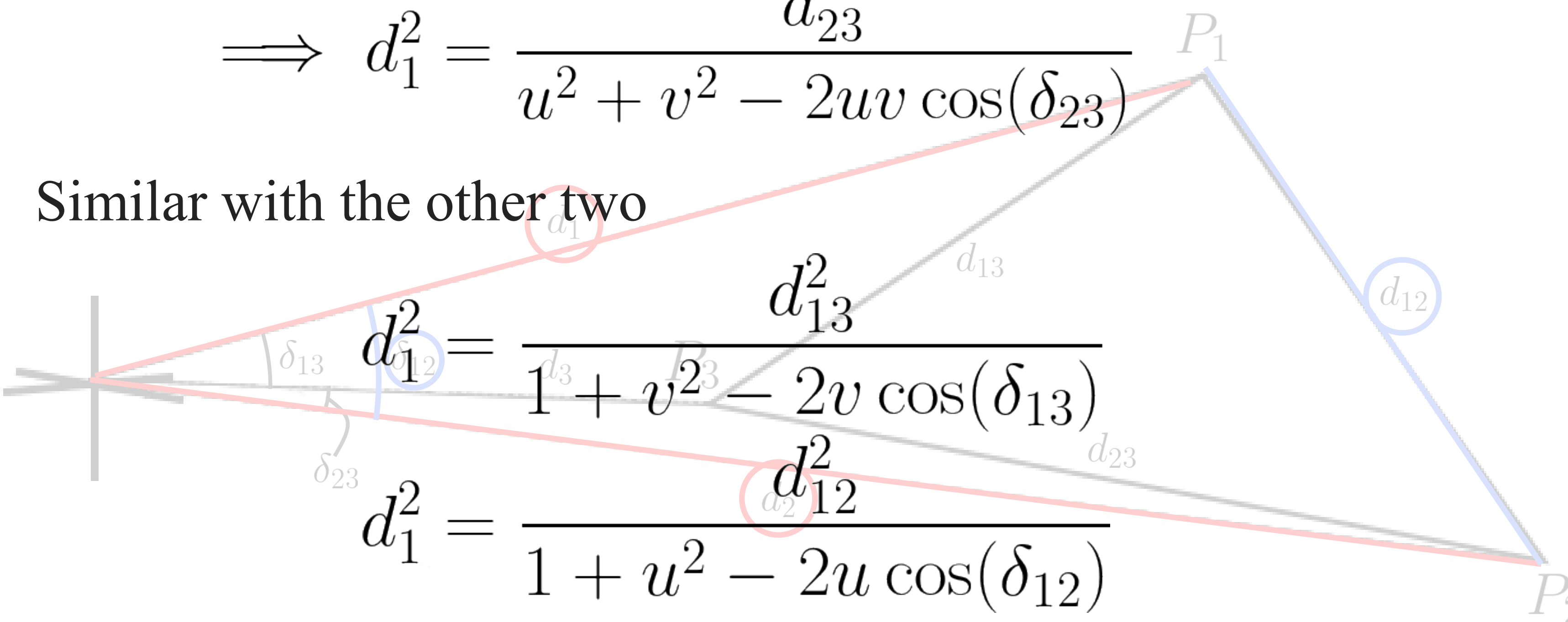
$$\implies u^2d_1^2 + v^2d_1^2 - d_1^2 2uv\cos(\delta_{23}) = d_{23}^2$$

$$\implies d_1^2 = \frac{d_{23}^2}{u^2 + v^2 - 2uv\cos(\delta_{23})}$$

Similar with the other two

$$d_1^2 = \frac{d_{13}^2}{1 + v^2 - 2v\cos(\delta_{13})}$$

$$d_1^2 = \frac{d_{12}^2}{1 + u^2 - 2u\cos(\delta_{12})}$$



Solving P3P

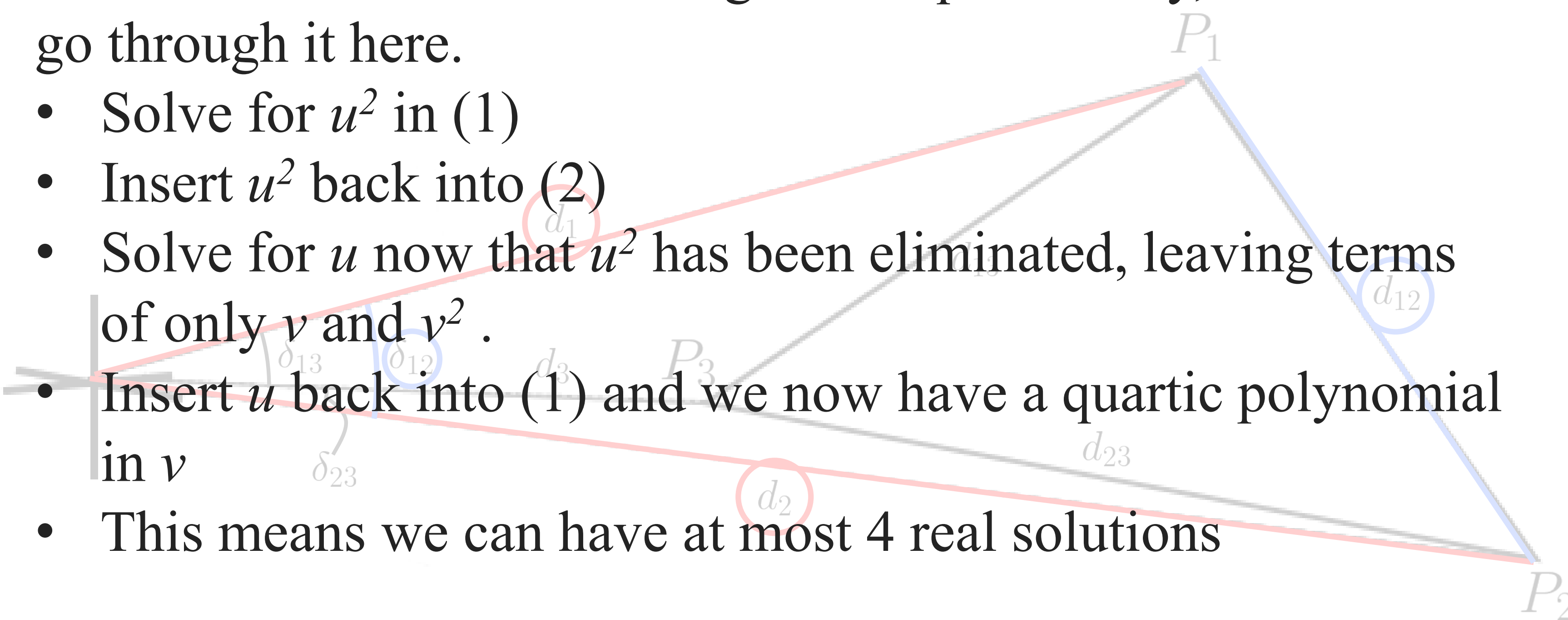
This reduces it to 2 quadratic equations in *u* and *v*:

$$d_{13}^2(u^2 + v^2 - 2uv\cos(\delta_{23})) = d_{23}^2(1 + v^2 - 2v\cos(\delta_{13})) \quad (1)$$

$$d_{12}^2(1 + v^2 - 2v\cos(\delta_{13})) = d_{13}^2(1 + u^2 - 2u\cos(\delta_{12})) \quad (2)$$

Now even more tricks. The algebra is quite messy, so we won't go through it here.

- Solve for u^2 in (1)
- Insert u^2 back into (2)
- Solve for *u* now that u^2 has been eliminated, leaving terms of only *v* and v^2 .
- Insert *u* back into (1) and we now have a quartic polynomial in *v*
- This means we can have at most 4 real solutions



Complete solution classification for the perspective-three-point problem
http://www.mmrc.iss.ac.cn/~xgao/paper/ieee.pdf

Next Lecture

Now that we have a way to solve for the distances, we still need to find the rotation and translation

$$d_i \begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix} = R \begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix} + T$$

So we are matching 3D point to 3D point – this is known as the Procrustes problem