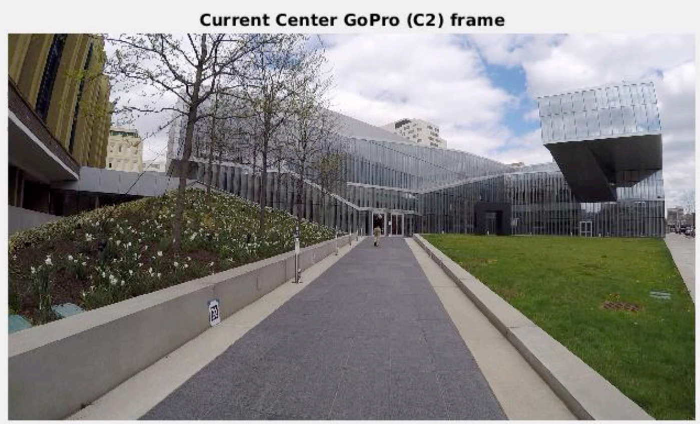
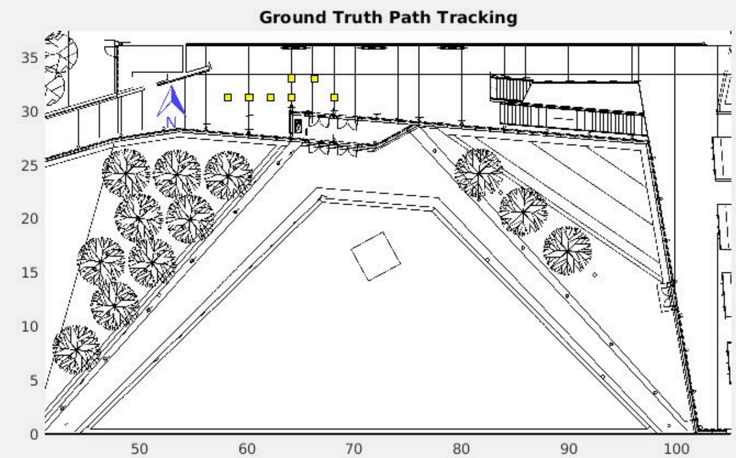
3D Motion from Two Views or Structure from Motion (SfM)

Kostas Daniilidis

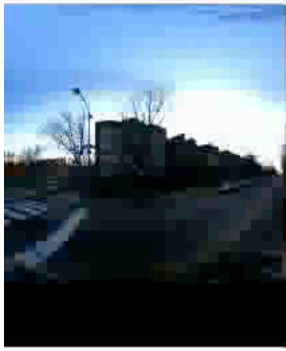
Main ingredient of all visual odometry algorithms



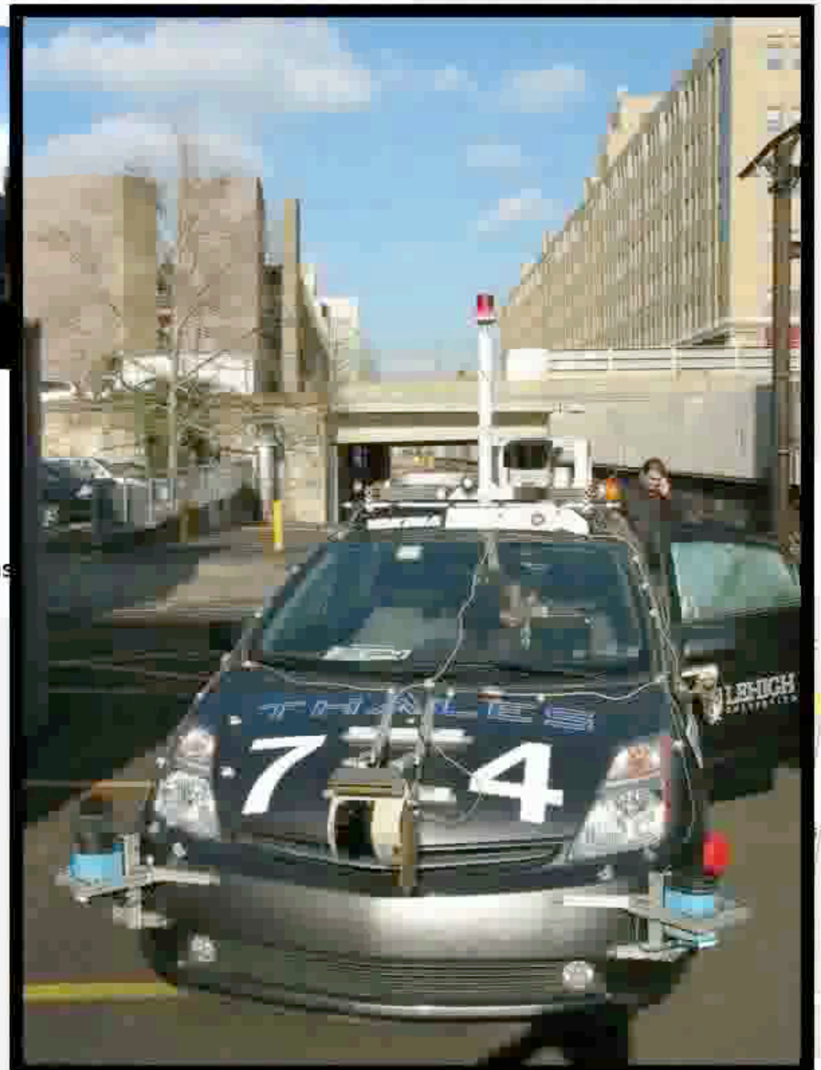


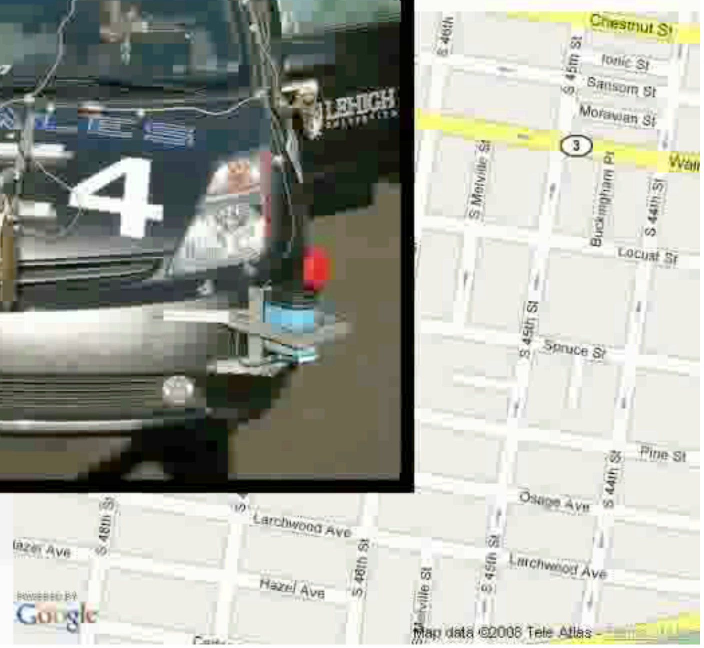
Panoramic image (from 6 cameras)

Reconstruction (global view)



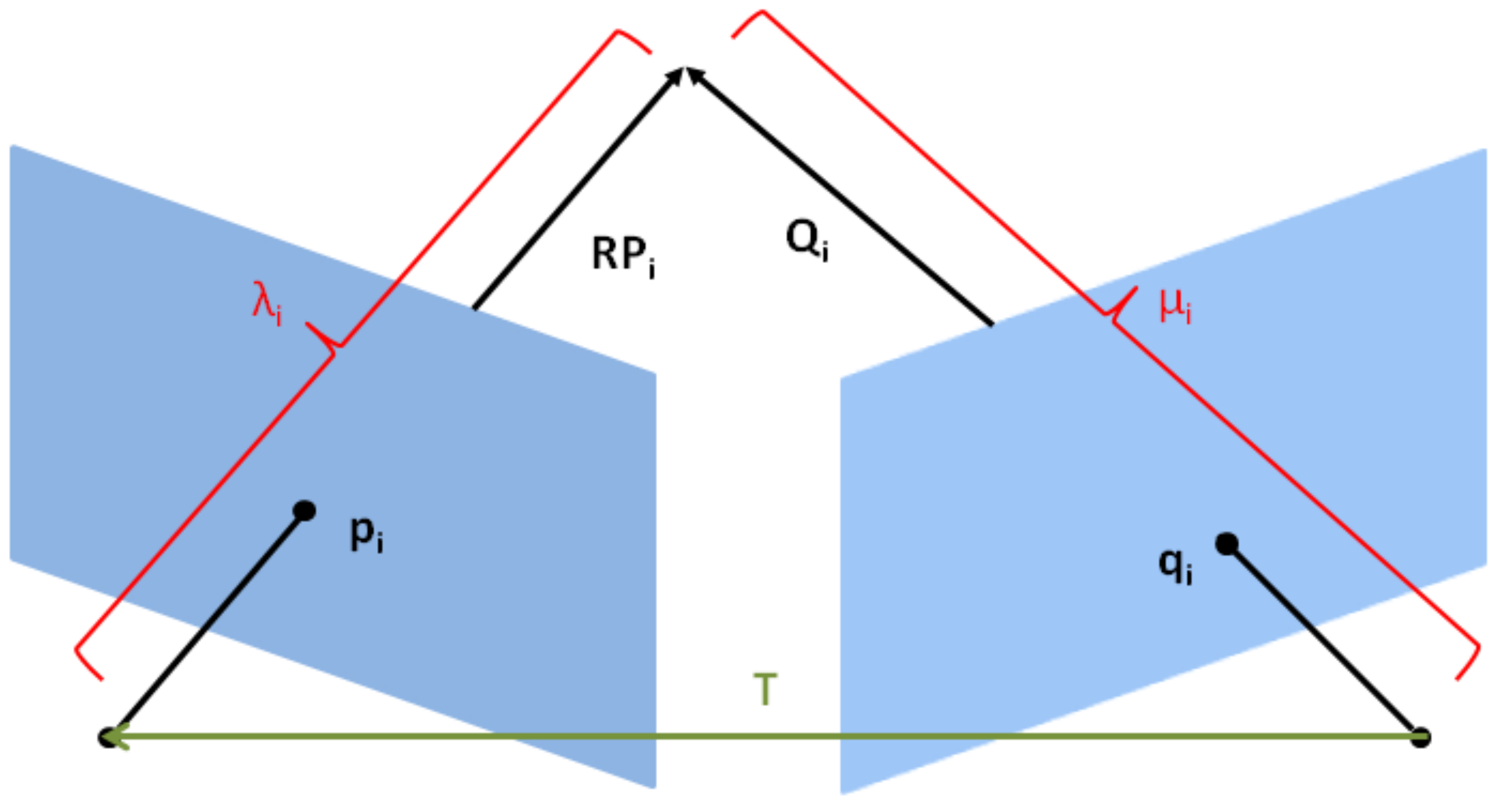
Recons





Two calibrated views of a scene

Two calibrated views of a scene

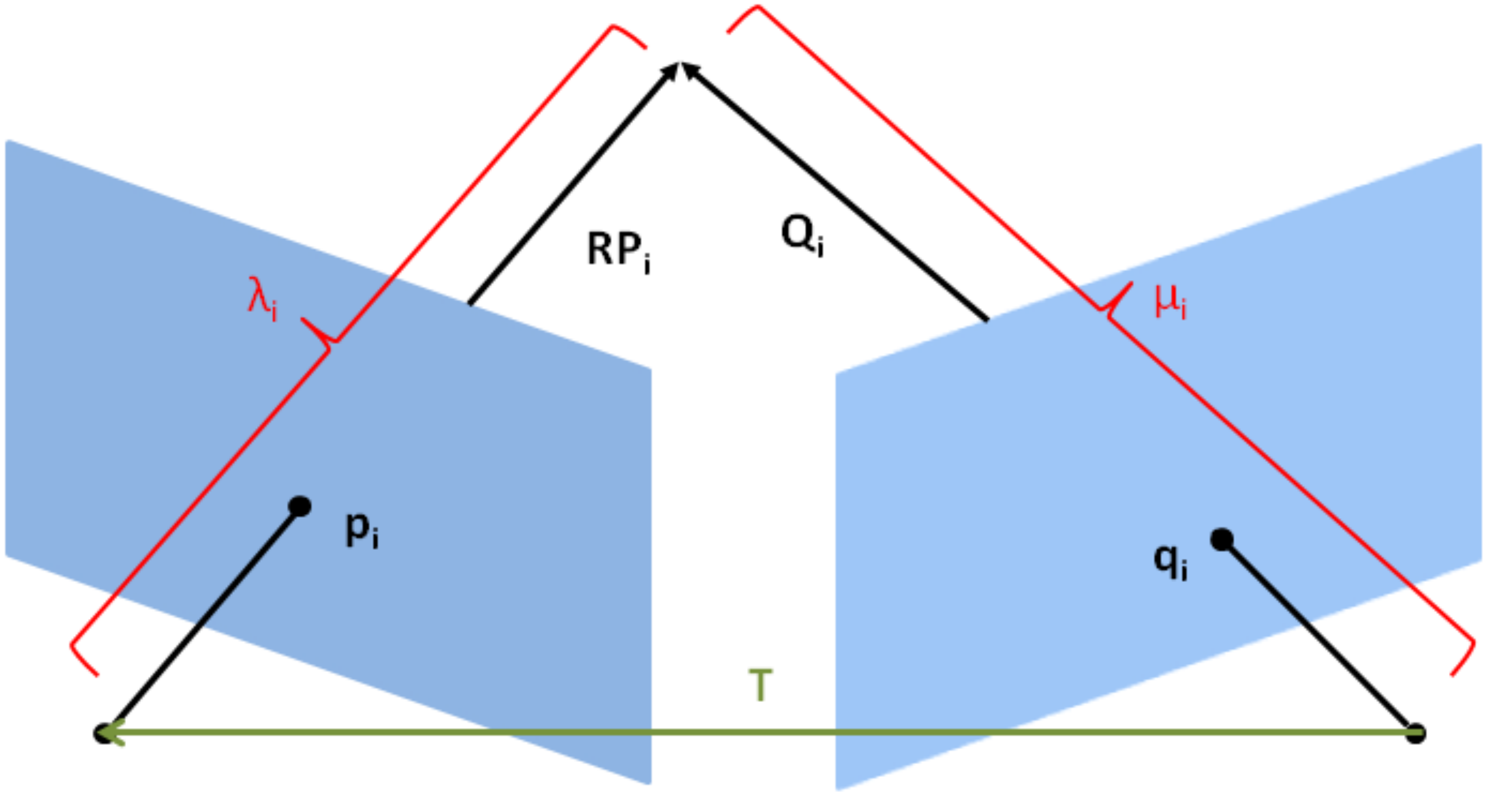


$$\lambda q = R\mu p + T$$

Given 2D correspondences (p,q)

Find motion R, T and depths λ, μ.

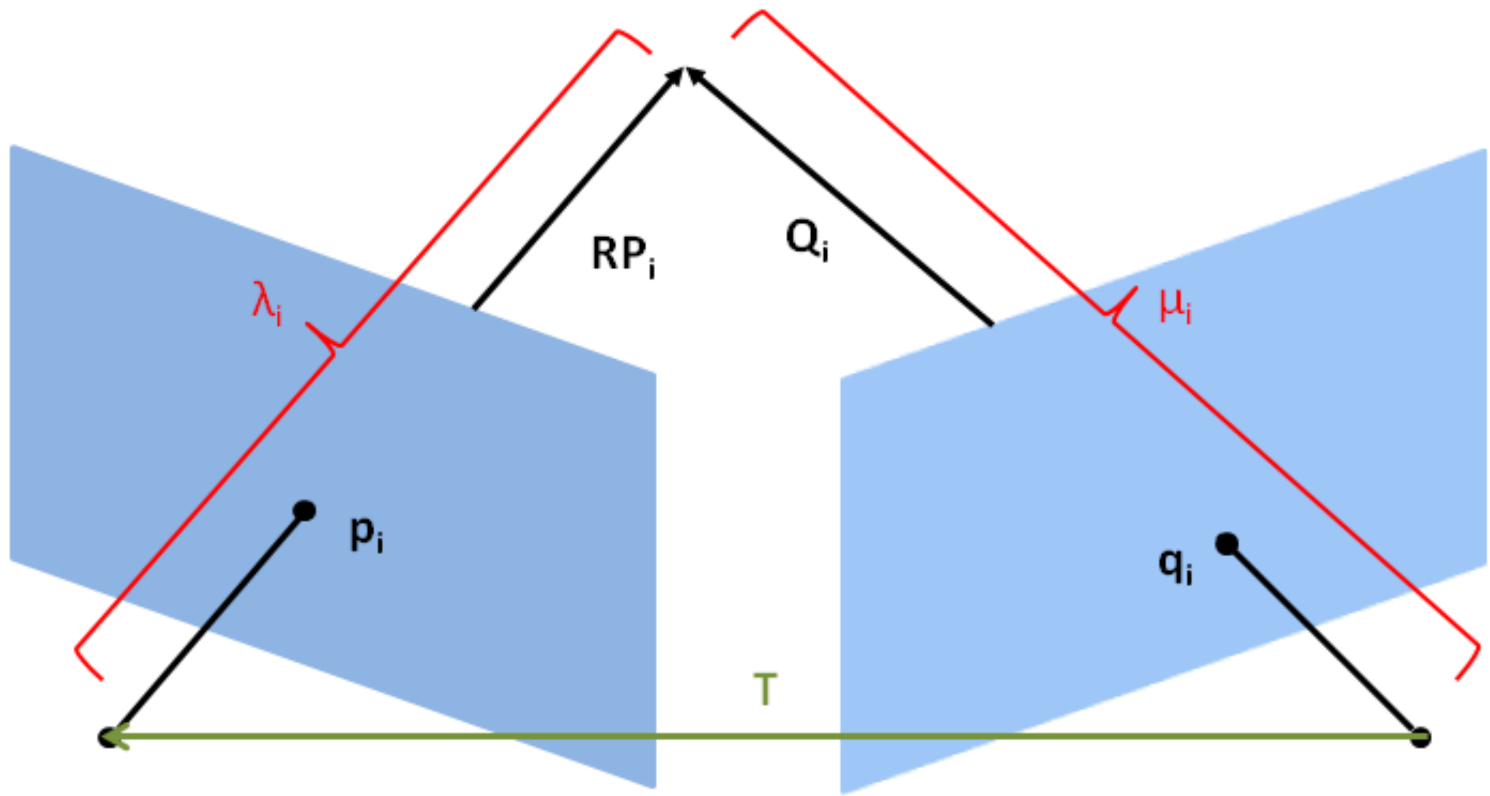
Epipolar constraint



We can eliminate the depths from $\lambda q = R\mu p + T$ and obtain the epipolar constraint:

$$q^T(T \times Rp) = 0$$

Geometric meaning of the epipolar constraint

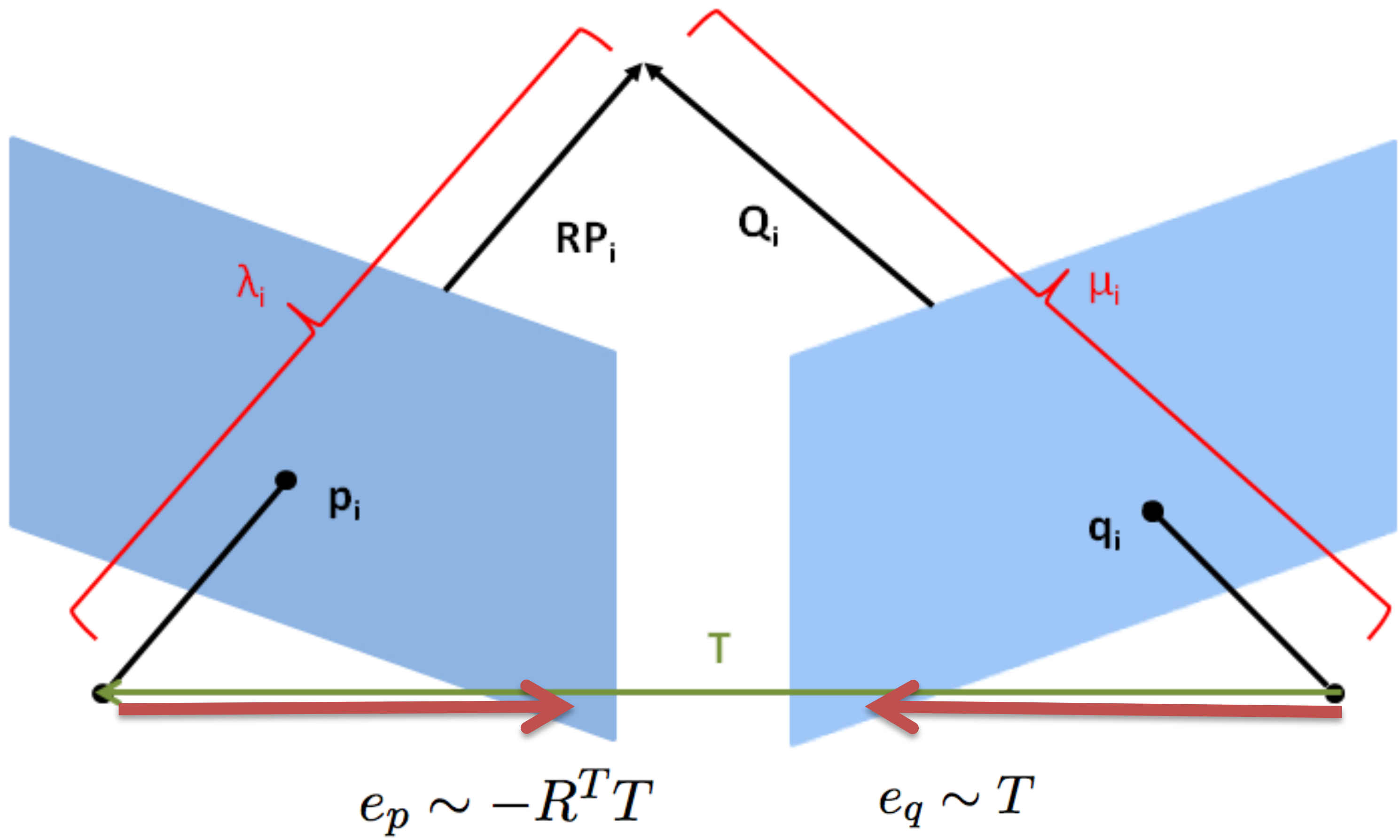


The two rays q and Rp intersect in space if and only if they are coplanar with the translation vector T. Three vectors are coplanar if their mixed product vanishes:

$$q^T(T \times Rp) = 0$$

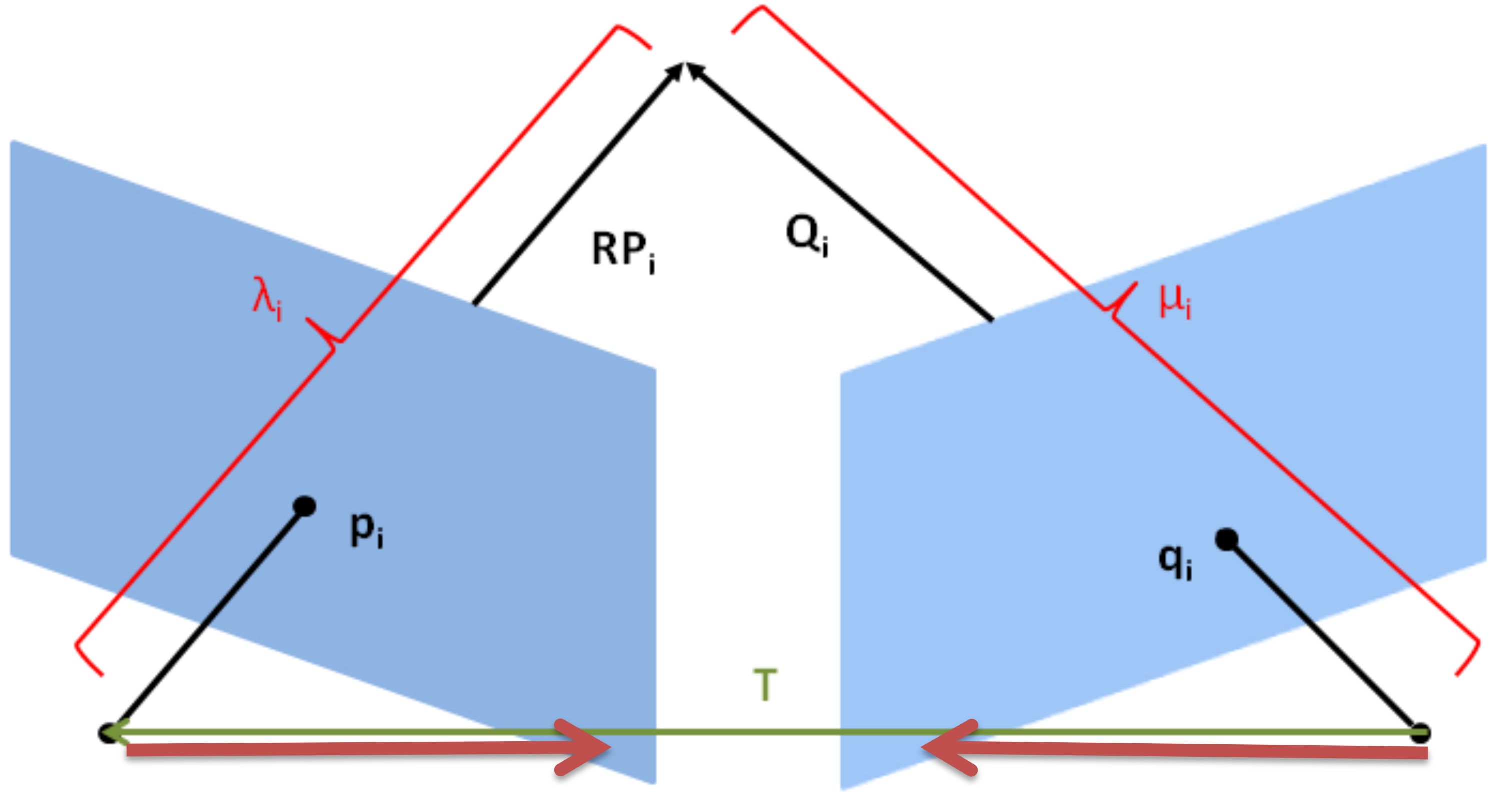
The plane spanned by the three vectors is called **epipolar** plane.

Why is it called epipolar?



$e_p \sim -R^T T$ and $e_q \sim T$ are the intersections of the baseline (translation) with the two image planes.

Essential matrix

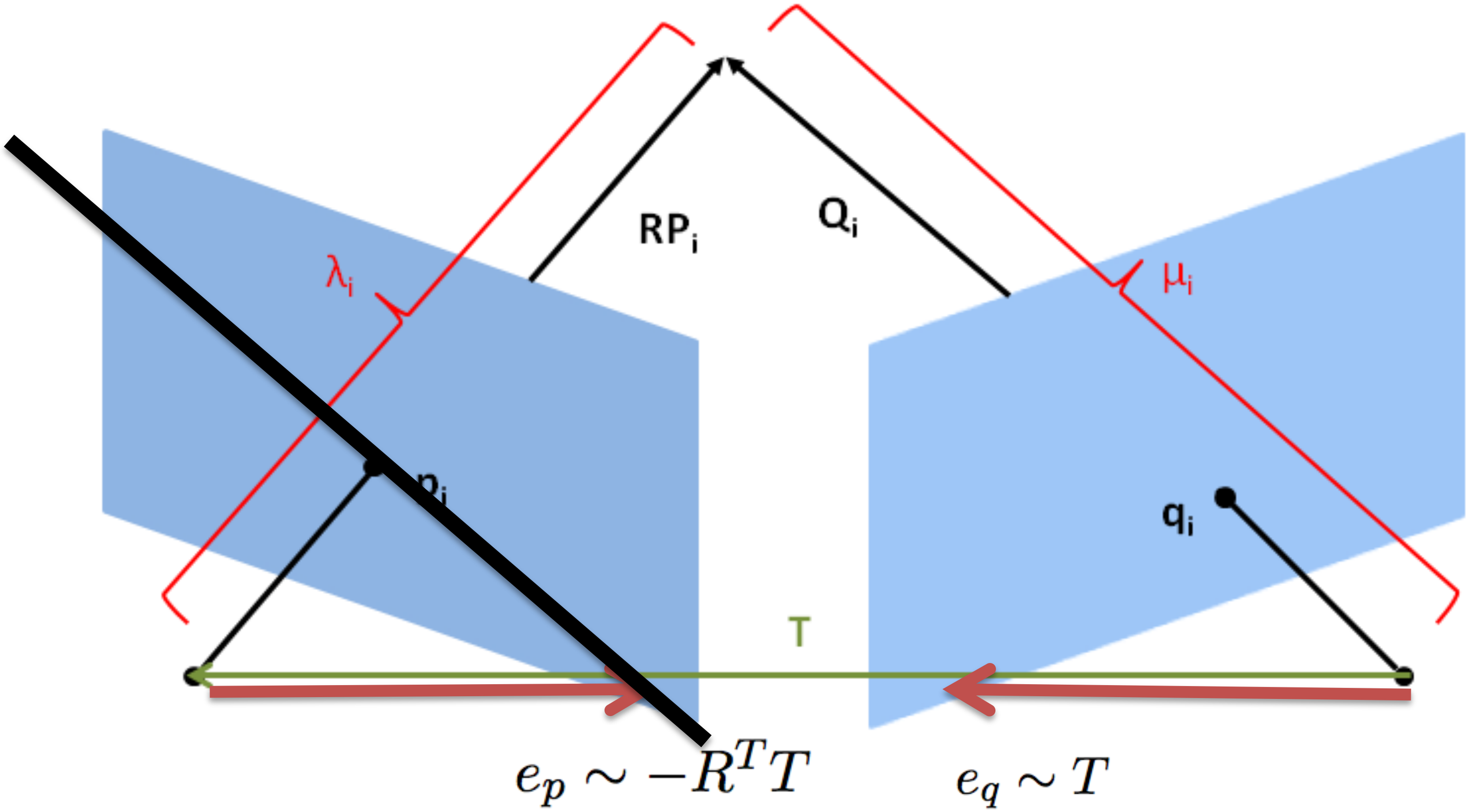


$$e_p \sim -R^T T \qquad e_q \sim T$$

We can summarize the unknowns into a 3x3 matrix E we will call the **essential** matrix:

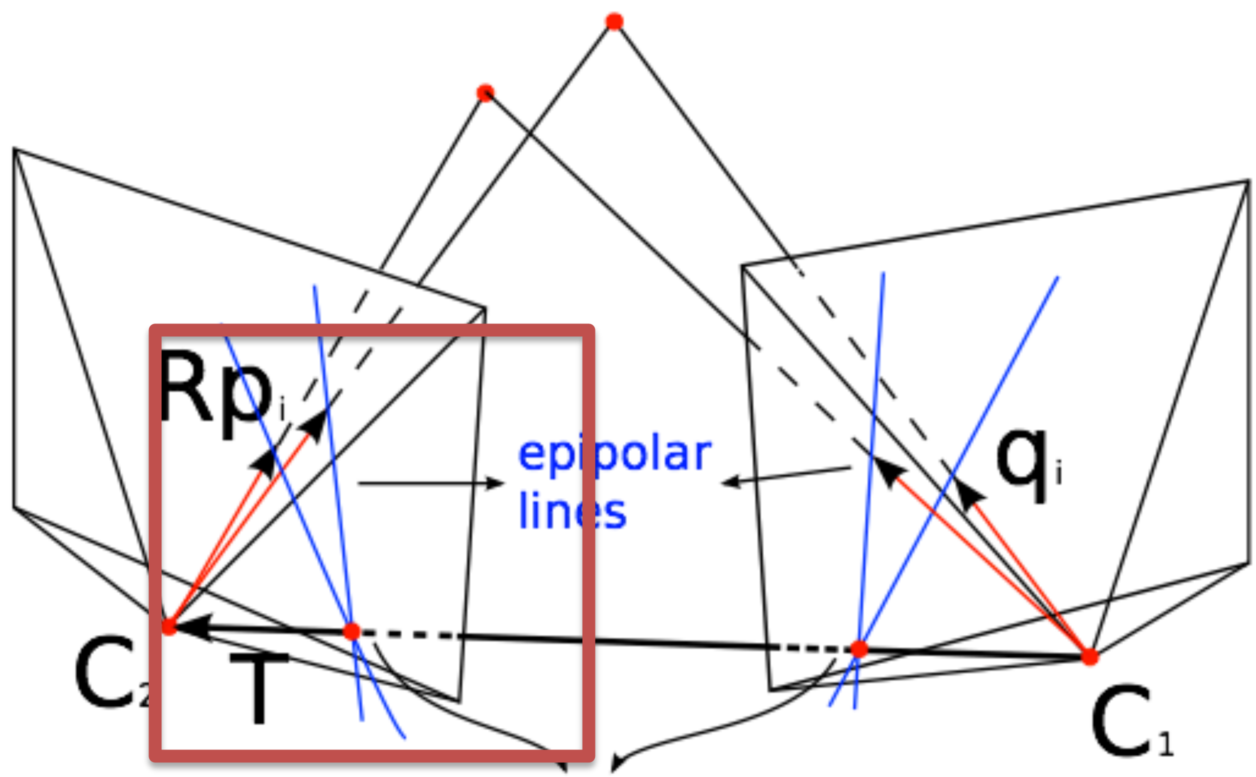
$$q^T E p = 0 \qquad \text{where} \qquad E = \hat{T} R$$

Geometric properties



Equation $q^T E p = 0$ is a line equation in the p-plane with line coefficients $E^T q$. It is called the epipolar line in p-plane.

Epipolar pencil of lines

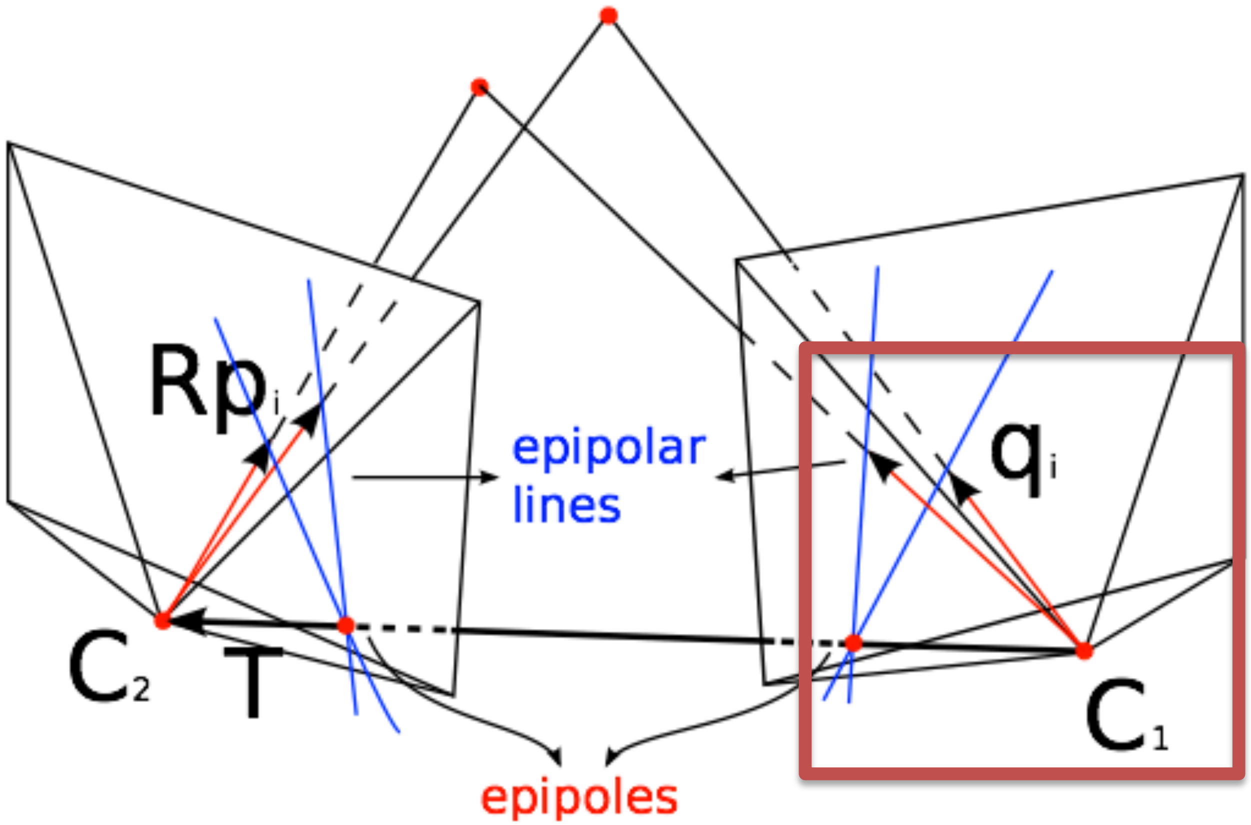


All epipolar lines $q^T E p = 0$ in p-plane go through the epipole e_p because

$$Ee_p = \hat{T}R(-R^T T) = \hat{T}T = T \times T = 0$$

They build a pencil of epipolar lines with one line per point correspondence.

Epipolar pencil of lines

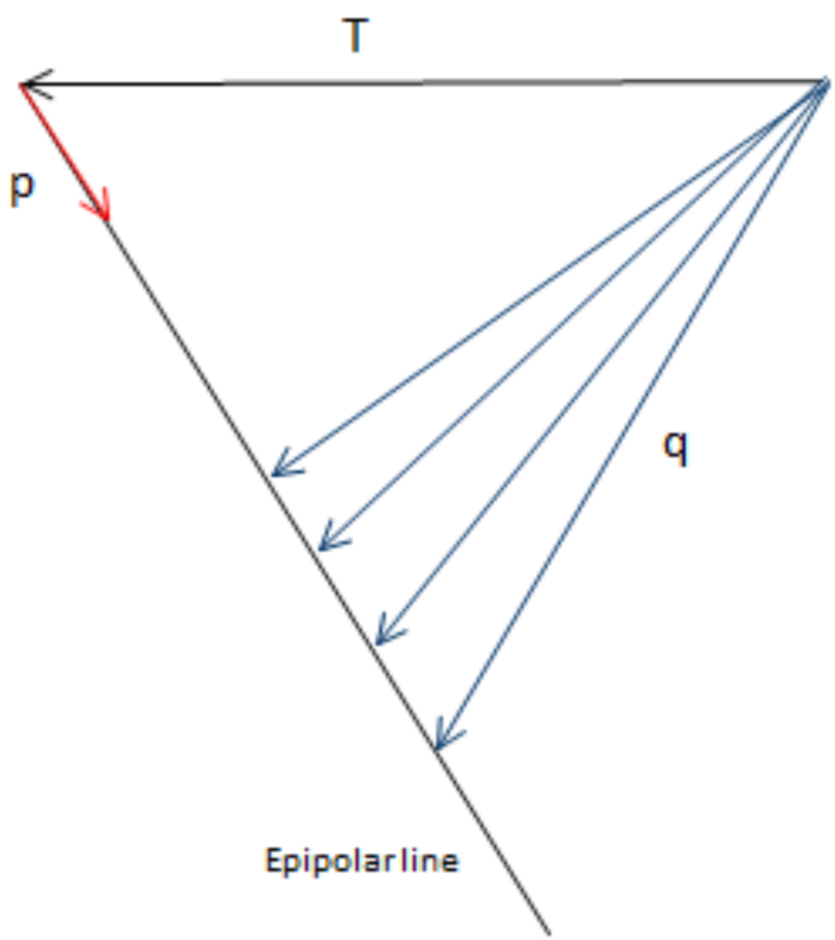


All epipolar lines $p^T E^T q = 0$ in the q-plane go through the epipole e_q because

$$E^T e_p = R^T \hat{T}^T T = R(\hat{T} T) = R(T \times T) = 0$$

They also build a pencil of epipolar lines with one line per point correspondence.

Simpler correspondence



Knowledge of the E-matrix allows us to search for points q corresponding to points p along the epipolar line, reducing correspondence to $1D$-search.

Position of the corresponding point q along epipolar line varies with depth of the 3D points which is still constrained to lie on the ray through p.

How can we compute the E-matrix?

If

$$E = \begin{pmatrix} e_1 & e_2 & e_3 \end{pmatrix}$$

then epipolar constraint can be rewritten as

$$\begin{aligned} q^T \begin{pmatrix} e_1 & e_2 & e_3 \end{pmatrix} \begin{pmatrix} p_x \\ p_y \\ p_z \end{pmatrix} &= q^T \left(p_x e_1 + p_y e_2 + p_z e_3 \right) \\ &= \begin{pmatrix} p_x q^T & p_y q^T & p_z q^T \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix} = 0 \end{aligned}$$

This equation is linear and homogeneous in $E' = \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix}$.

The 8-point algorithm

Let $\vec{a} = \begin{pmatrix} p_x q^T & p_y q^T & p_z q^T \end{pmatrix}$

$$\begin{pmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_n^T \end{pmatrix} E' = 0$$

where a_i is the known 1 x 9 vector of image points and E' is the essential matrix re-organized into a 9 x 1 column vector.

E' has to be in the null-space of $\begin{pmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_n^T \end{pmatrix}$.

Properties of the Essential matrix

$$E^T = R^T \hat{T} T = 0$$

E is a singular matrix, $\det(E) = 0$.

$$\begin{aligned} EE^T &= \widehat{T}\widehat{T}^T \\ &= TT^T - T^T T I \\ &= \begin{bmatrix} t_x^2 & t_x t_y & t_x t_z \\ t_x t_y & t_y^2 & t_y t_z \\ t_x t_z & t_y t_z & t_z^2 \end{bmatrix} - \|T\|^2 I \end{aligned}$$

Properties of the Essential matrix

$$
\begin{aligned}
EE^T &= \widehat{T}\widehat{T}^T \\
&= TT^T - T^T T I \\
&= \begin{bmatrix} t_x^2 & t_x t_y & t_x t_z \\ t_x t_y & t_y^2 & t_y t_z \\ t_x t_z & t_y t_z & t_z^2 \end{bmatrix} - \|T\|^2 I
\end{aligned}
$$

If we solve the characteristic polynomial $\det(EE^T - \lambda I) = 0$ we will find two eigenvalues both equal to $\|T\|^2$.

Properties of the Essential matrix

Recall that the singular values of E are the square-roots of the eigenvalues of EE^T if E is a square matrix.

Hence, we have proved that **if a matrix is essential, namely, can be decomposed as the product of an antisymmetric $\hat{T}$ and a special orthogonal R then its singular values are $\sigma_1 = \sigma_2 > 0$ and $\sigma_3 = 0$.**

This is helpful in checking of a matrix is essential but is not constructive on how to decompose it.

Properties of the Essential matrix

We have to prove the sufficient condition:

If the singular values of a matrix are are $\sigma_1 = \sigma_2 > 0$ and $\sigma_3 = 0$ then the matrix can be decomposed into the product of an antisymmetric $\hat{T}$ and a special orthogonal R.

We need a lemma !

If Q is orthogonal ($Q^TQ = I$), then

$$\widehat{Qa} = Q\hat{a}Q^T$$

Proof: $\widehat{Qa}b = Qa \times b = Q(a \times Q^Tb) = Q\hat{a}Q^Tb.$

and the following simple fact

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = T_z^T R_{z,\pi/2}$$

We need a lemma !

If Q is orthogonal ($Q^T Q = I$), then

$$\widehat{Qa} = Q\hat{a}Q^T$$

Proof: $\widehat{Qa}b = Qa \times b = Q(a \times Q^T b) = Q\hat{a}Q^T b.$

and the following simple fact

$$\boxed{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = T_z^T R_{z,\pi/2}}$$

$$
\begin{aligned}
E = &\quad U \begin{bmatrix} \sigma & 0 & 0 \\ 0 & \sigma & 0 \\ 0 & 0 & 0 \end{bmatrix} V^T \\
= &\quad \sigma U \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} V^T \\
= &\quad \sigma U \widehat{T_z}^T R_z V^T \\
= &\quad \sigma \underbrace{\widehat{UT_z}}_{\text{antisymmetric}} \underbrace{URV^T}_{\text{orthogonal}}
\end{aligned}
$$

Observe $UT_z = U \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$, which is the last column of U

Necessary and sufficient condition: E is essential **iff** $\sigma_1(E) = \sigma_2(E) \neq 0$ and $\sigma_3(E) = 0$.

We just showed that there is at least one such decomposition $\hat{T}R$, but is it unique?

We showed the following decomposition:

$$\boxed{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{-\widehat{T_z}} \underbrace{\begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{R_{z,\pi/2}}}$$

But we could similarly write $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \widehat{T_z} R_{z,-\pi/2}$.

We just showed that there is at least one such decomposition $\hat{T}R$, but is it unique?

We showed the following decomposition:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{-\widehat{T_z}} \underbrace{\begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{R_{z,\pi/2}}$$

But we could similarly write $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \widehat{T_z} R_{z,-\pi/2}$.

If $E = U\Sigma V^T = U \begin{bmatrix} \sigma & 0 & 0 \\ 0 & \sigma & 0 \\ 0 & 0 & 0 \end{bmatrix} V^T$, there are four solutions for the pair $(\widehat{T}, R)$:

$$
\begin{aligned}
(\widehat{T_1}, R_1) = &\quad (UR_{z,+\pi/2}\Sigma U^T, UR^T_{z,+\pi/2}V^T) \\
(\widehat{T_2}, R_2) = &\quad (UR_{z,-\pi/2}\Sigma U^T, UR^T_{z,-\pi/2}V^T) \\
(\widehat{T_1}, R_2) = &\quad (UR_{z,+\pi/2}\Sigma U^T, UR^T_{z,-\pi/2}V^T) \\
(\widehat{T_2}, R_1) = &\quad (UR_{z,-\pi/2}\Sigma U^T, UR^T_{z,+\pi/2}V^T)
\end{aligned}
$$

Please remember that we have to make R have $\det(R) = 0$, see our Procrustes lecture.

Mirror ambiguity: If T is a solution, then $-T$ is a solution, too. There is no way to disambiguate from the epipolar constraint: $q^T(-T \times Rp) = 0$.

Twisted pair ambiguity: If R is a solution, then also $R_{T,\pi}R$ is a solution. The first image is "twisted" around the baseline 180 degrees.

Mirror ambiguity: If T is a solution, then $-T$ is a solution, too. There is no way to disambiguate from the epipolar constraint: $q^T(-T \times Rp) = 0$.

Twisted pair ambiguity: If R is a solution, then also $R_{T,\pi}R$ is a solution. The first image is "twisted" around the baseline 180 degrees.

The full two-view algorithm

1. Build the homogeneous linear system by stacking epipolar constraints $q_i^T(T \times Rp_i) = 0,\ i = 1, \ldots, 8$:

$$\underbrace{\begin{bmatrix} \vdots \\ (q_i \otimes p_i)^T \\ \vdots \end{bmatrix}}_{A\ (8 \times 9)} \begin{bmatrix} e_1' \\ e_2' \\ e_3' \end{bmatrix}$$

2. Let $\begin{bmatrix} e_1' \\ e_2' \\ e_3' \end{bmatrix}$ be the nullspace of A (if $\sigma_8 \approx 0$ give up)

The full two-view algorithm

1. Build the homogeneous linear system by stacking epipolar constraints $q_i^T(T \times Rp_i) = 0,\ i = 1, \ldots, 8$:

$$\underbrace{\begin{bmatrix} \vdots \\ (q_i \otimes p_i)^T \\ \vdots \end{bmatrix}}_{A\ (8\times 9)} \begin{bmatrix} e_1' \\ e_2' \\ e_3' \end{bmatrix}$$

2. Let $\begin{bmatrix} e_1' \\ e_2' \\ e_3' \end{bmatrix}$ be the nullspace of A (if $\sigma_8 \approx 0$ give up)

The full two-view algorithm

3. $\begin{bmatrix} e'_1 & e'_2 & e'_3 \end{bmatrix} = U\text{diag}\,(\sigma'_1, s'_2, \sigma'_3)V^T$. Then use the following estimate of the essential matrix:

$$E = U\text{diag}\,\left(\frac{\sigma'_1 + \sigma'_2}{2}, \frac{\sigma'_1 + \sigma'_2}{2}, 0\right) V^T$$

4. $T = \pm\hat{u}_3 \quad R = UR_{Z,\pi/2}V^T$ or $R = R_{T,\pi}R$
5. Try all four pairs (T, R) to check if reconstructed points are **in front** of the cameras $\boxed{\lambda q = \mu R p + T}$ give $\lambda, \mu > 0$.

The full two-view algorithm

3. $\begin{bmatrix} e'_1 & e'_2 & e'_3 \end{bmatrix} = U\text{diag}\,(\sigma'_1, s'_2, \sigma'_3)V^T$. Then use the following estimate of the essential matrix:

$$E = U\text{diag}\left(\frac{\sigma'_1+\sigma'_2}{2}, \frac{\sigma'_1+\sigma'_2}{2}, 0\right)V^T$$

4. $T = \pm\hat{u}_3 \quad R = UR_{Z,\pi/2}V^T$ or $R = R_{T,\pi}R$
5. Try all four pairs (T, R) to check if reconstructed points are **in front** of the cameras $\boxed{\lambda q = \mu Rp + T}$ give $\lambda, \mu > 0$.

Triangulation is possible if we have computed R and T but again up to a scale factor. Set $\|T\| = 1$:

$$\underbrace{(q_i - Rp_i)}_{3\times 2}\underbrace{\begin{pmatrix}\mu_i \\ \lambda_i\end{pmatrix}}_{2\times 1} = \underbrace{T}_{3\times 1}$$

There are then 3 equations with 2 unknowns λ_i and μ_i for each point.

Solve with pseudo-inverse.

Structure and motion from two views