

Two-View Geometry

Sami Osborn

2025

1 Overview

This document develops the mathematical foundations of two-view geometry.

Given two images of a static scene, our goals are:

- Estimate the fundamental matrix
- Recover the relative camera pose (up to scale)
- Triangulate 3D points

2 Single-View Camera Model

A pinhole camera maps 3D points in the world to 2D image coordinates via a projective transformation.

2.1 Camera Projection Matrix

Let $X = (X, Y, Z, 1)^\top \in \mathbb{P}^3$ be a 3D point in homogeneous coordinates.

A camera is represented by the projection matrix

$$P \in \mathbb{R}^{3 \times 4}, \quad x \sim PX, \quad x \in \mathbb{P}^2$$

Definition 2.1 (Intrinsics and extrinsics). A common parameterisation is

$$P = K [R \mid t]$$

where $K \in \mathbb{R}^{3 \times 3}$ is the intrinsic calibration matrix, and (R, t) describes the rigid transform between world and camera frames with $R \in SO(3)$, $t \in \mathbb{R}^3$

2.2 Intrinsic Matrix

The intrinsic calibration matrix $K \in \mathbb{R}^{3 \times 3}$ maps camera-frame coordinates to pixel coordinates:

$$K = \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

Given \hat{x} in intrinsic normalised (calibrated) image coordinates, pixel coordinates are $x = K\hat{x}$.

Conversely, $\hat{x} = K^{-1}x$.

2.3 Rotation and Translation

The rotation matrix $R \in SO(3)$ satisfies

$$R^\top R = I, \quad \det(R) = 1$$

This maps vectors from world coordinates into camera coordinates.

The translation vector $t \in \mathbb{R}^3$ determines the camera centre position relative to the world frame (in this convention, t is expressed in the camera frame).

3 Two Views and Relative Pose

Consider two cameras observing the same 3D point X :

$$x_1 \sim P_1 X, \quad x_2 \sim P_2 X$$

With

$$P_i = K_i[R_i \mid t_i], \quad i = 1, 2$$

3.1 Relative Pose Simplification

Proposition 3.1 (Choice of world frame). *WLOG, one may choose world coordinates so that the first camera is at the origin with identity orientation, i.e. for some relative pose (R, t)*

$$P_1 = K_1[I \mid 0], \quad P_2 = K_2[R \mid t]$$

Proof. Trivial. □

4 Epipolar Geometry

The two camera centres and a 3D point X define the epipolar plane.

Its intersection with each image plane is an epipolar line.

4.1 A key algebraic tool

Definition 4.1 (Cross-product matrix). For $t = (t_x, t_y, t_z)^\top$, define the skew-symmetric matrix

$$[t]_\times = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix}$$

so that for any $a \in \mathbb{R}^3$, $[t]_\times a = t \times a$

5 The Essential Matrix (Calibrated Case)

Assume both cameras are calibrated, and work with intrinsic normalised image coordinates

$$\hat{x}_1 = K_1^{-1}x_1, \quad \hat{x}_2 = K_2^{-1}x_2$$

Theorem 5.1 (Essential matrix and epipolar constraint). *Let $P_1 = [I \mid 0]$, $P_2 = [R \mid t]$ be the normalised projection matrices (intrinsic removed). Then, there exists a matrix $E \in \mathbb{R}^{3 \times 3}$ such that for corresponding normalised points \hat{x}_1, \hat{x}_2 ,*

$$\hat{x}_2^\top E \hat{x}_1 = 0, \quad \text{and} \quad E = [t]_\times R$$

Proof. Let the 3D point in the first camera frame be $X_1 \in \mathbb{R}^3$

Then the second camera frame sees

$$X_2 = RX_1 + t$$

Since \hat{x}_1 and \hat{x}_2 represent (homogeneous) directions of the viewing rays, \exists scalars $\lambda_1, \lambda_2 \neq 0$ such that

$$X_1 = \lambda_1 \hat{x}_1, \quad X_2 = \lambda_2 \hat{x}_2$$

Hence

$$\lambda_2 \hat{x}_2 = R(\lambda_1 \hat{x}_1) + t$$

Rearrange to isolate t :

$$t = \lambda_2 \hat{x}_2 - \lambda_1 R \hat{x}_1$$

The vectors \hat{x}_2 , $R\hat{x}_1$, and t lie in the epipolar plane, so they are coplanar and therefore

$$\hat{x}_2^\top [t]_\times (R\hat{x}_1) = 0$$

Thus $\hat{x}_2^\top E \hat{x}_1 = 0$ with $E := [t]_\times R$ □

5.1 Algebraic properties

Proposition 5.2 (Rank constraint). *If $t \neq 0$, then $\text{rank}(E) = 2$.*

Proof. Since R is invertible, $\text{rank}([t]_\times R) = \text{rank}([t]_\times)$. For nonzero t , the linear map $a \mapsto t \times a$ has a 1D nullspace which is $\text{span}\{t\}$, so $\text{rank}([t]_\times) = 2$, and $\text{rank}(E) = 2$. □

Proposition 5.3 (Singular values of E). *Any essential matrix $E = [t]_\times R$ has two equal nonzero singular values and one zero singular value.*

Proof. Write $E = [t]_\times R$. Since R is orthogonal,

$$EE^\top = [t]_\times R R^\top [t]_\times^\top = [t]_\times [t]_\times^\top.$$

A standard identity gives

$$[t]_\times [t]_\times^\top = \|t\|^2 I - tt^\top.$$

The matrix $\|t\|^2 I - tt^\top$ has eigenvalue 0 in direction t , and eigenvalue $\|t\|^2$ on the two-dimensional subspace orthogonal to t .

Hence EE^\top has eigenvalues $\|t\|^2, \|t\|^2, 0$, so E has singular values $\|t\|, \|t\|, 0$. □

6 The Fundamental Matrix (Pixel Coordinates)

For pixel coordinates $x_i = K_i \hat{x}_i$

Theorem 6.1 (Relation between F and E). *The fundamental matrix F is the unique (up to scale) matrix satisfying*

$$x_2^\top F x_1 = 0 \quad \text{for all corresponding pixels } (x_1, x_2)$$

and is related to the essential matrix by

$$E = K_2^\top F K_1, \quad \text{or} \quad F = K_2^{-\top} E K_1^{-1}$$

Proof. Start from the essential constraint in intrinsic normalised coordinates:

$$\hat{x}_2^\top E \hat{x}_1 = 0$$

Substitute $\hat{x}_1 = K_1^{-1} x_1$, $\hat{x}_2 = K_2^{-1} x_2$:

$$(K_2^{-1} x_2)^\top E (K_1^{-1} x_1) = 0 \iff x_2^\top (K_2^{-\top} E K_1^{-1}) x_1 = 0$$

Thus, $F := K_2^{-\top} E K_1^{-1}$ satisfies the pixel epipolar constraint. Rearranging gives $E = K_2^\top F K_1$ \square

6.1 Geometric meaning and basic properties

Proposition 6.2 (Epipolar lines). *Given $x_1 \in \mathbb{P}^2$, define $l_2 := F x_1$. Then any corresponding point x_2 satisfies $x_2^\top l_2 = 0$, i.e. x_2 lies on the line l_2 . Similarly, $l_1 := F^\top x_2$ is the epipolar line in image 1.*

Proof. The epipolar constraint is $x_2^\top F x_1 = 0$. Set $l_2 := F x_1$. Then $x_2^\top l_2 = 0$, which is precisely the incidence relation “point x_2 lies on line l_2 ” in homogeneous coordinates. \square

Proposition 6.3 (Rank-2 constraint). $\text{rank}(F) = 2$.

Proof. From $F = K_2^{-\top} E K_1^{-1}$, and since K_i are invertible, $\text{rank}(F) = \text{rank}(E)$. In the non-degenerate case $t \neq 0$, $\text{rank}(E) = 2$ by the earlier proposition, hence $\text{rank}(F) = 2$. \square

Proposition 6.4 (Epipoles as nullspaces). *There exist points $e_1, e_2 \in \mathbb{P}^2$ (the epipoles) such that*

$$F e_1 = 0, \quad F^\top e_2 = 0$$

Proof. Since $\text{rank}(F) = 2$, both nullspaces are one-dimensional. Choose any nonzero $e_1 \in \ker(F)$ and $e_2 \in \ker(F^\top)$. \square

7 Two-View Estimation from Point Correspondences

7.1 Linear Estimation of the Fundamental Matrix

Proposition 7.1 (Linear constraint from one correspondence). *Let $x_1 = (u_1, v_1, 1)^\top$ and $x_2 = (u_2, v_2, 1)^\top$. Writing $f = \text{vec}(F) \in \mathbb{R}^9$, the epipolar constraint $x_2^\top F x_1 = 0$ is equivalent to a homogeneous linear equation $a^\top f = 0$, where*

$$a = [u_2 u_1, u_2 v_1, u_2, v_2 u_1, v_2 v_1, v_2, u_1, v_1, 1]$$

Proof. Write $F = (f_{ij})$. Then

$$x_2^\top F x_1 = \begin{bmatrix} u_2 & v_2 & 1 \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} u_1 \\ v_1 \\ 1 \end{bmatrix}$$

Expanding produces

$$(u_2 u_1) f_{11} + (u_2 v_1) f_{12} + (u_2) f_{13} + (v_2 u_1) f_{21} + (v_2 v_1) f_{22} + (v_2) f_{23} + (u_1) f_{31} + (v_1) f_{32} + f_{33} = 0$$

This is precisely $a^\top f = 0$ when f is the vector of stacked entries of F . \square

Proposition 7.2 (The 8-point linear system). *Given n correspondences $\{(x_1^{(i)}, x_2^{(i)})\}_{i=1}^n$. Define $A \in \mathbb{R}^{n \times 9}$ by stacking the row vectors $a^{(i)\top}$ from the previous proposition. Then the epipolar constraints are equivalent to:*

$$A f = 0$$

Under noise, a least-squares estimate is obtained by choosing f as the right singular vector of A corresponding to the smallest singular value.

Proof. By the variational characterisation of singular values (Rayleigh–Ritz), the minimiser is the right singular vector corresponding to the smallest singular value of A . \square

7.2 Hartley Normalisation

Theorem 7.3 (Hartley normalisation transform). *Let $\{x^{(i)}\}_{i=1}^n$ be pixel points with $x^{(i)} = (u^{(i)}, v^{(i)}, 1)^\top$. Define the centroid*

$$c_x := \frac{1}{n} \sum_{i=1}^n u^{(i)}, \quad c_y := \frac{1}{n} \sum_{i=1}^n v^{(i)}$$

Let translated coordinates be

$$\bar{u}^{(i)} := u^{(i)} - c_x, \quad \bar{v}^{(i)} := v^{(i)} - c_y$$

Define the mean Euclidean distance from the origin

$$d := \frac{1}{n} \sum_{i=1}^n \sqrt{(\bar{u}^{(i)})^2 + (\bar{v}^{(i)})^2}$$

Assuming $d > 0$, let $s := \sqrt{2}/d$ and define

$$T := \begin{bmatrix} s & 0 & -s c_x \\ 0 & s & -s c_y \\ 0 & 0 & 1 \end{bmatrix}$$

Then, the transformed points $\tilde{x}^{(i)} := T x^{(i)}$ satisfy:

1. *The centroid of the inhomogeneous coordinates of $\{\tilde{x}^{(i)}\}$ is at the origin*
2. *The average Euclidean distance of the inhomogeneous coordinates from the origin is $\sqrt{2}$*

Proof. Write $\tilde{x}^{(i)} = (\tilde{u}^{(i)}, \tilde{v}^{(i)}, 1)^\top$. By construction,

$$\tilde{u}^{(i)} = s(u^{(i)} - c_x) = s\bar{u}^{(i)}, \quad \tilde{v}^{(i)} = s(v^{(i)} - c_y) = s\bar{v}^{(i)}$$

For the centroid,

$$\frac{1}{n} \sum_i \tilde{u}^{(i)} = s \left(\frac{1}{n} \sum_i (u^{(i)} - c_x) \right) = s \left(\frac{1}{n} \sum_i u^{(i)} - c_x \right) = 0$$

Similarly, $\frac{1}{n} \sum_i \tilde{v}^{(i)} = 0$

For the average distance, note that

$$\sqrt{(\tilde{u}^{(i)})^2 + (\tilde{v}^{(i)})^2} = s \sqrt{(\bar{u}^{(i)})^2 + (\bar{v}^{(i)})^2}$$

Thus,

$$\frac{1}{n} \sum_i \sqrt{(\tilde{u}^{(i)})^2 + (\tilde{v}^{(i)})^2} = s \left(\frac{1}{n} \sum_i \sqrt{(\bar{u}^{(i)})^2 + (\bar{v}^{(i)})^2} \right) = sd = \sqrt{2}$$

□

Proposition 7.4 (Effect of normalisation on the fundamental matrix). *Let $\tilde{x}_1 = T_1 x_1$ and $\tilde{x}_2 = T_2 x_2$ be similarity transforms applied to the points in images 1 and 2.*

If \tilde{F} satisfies

$$\tilde{x}_2^\top \tilde{F} \tilde{x}_1 = 0 \quad \text{for all corresponding } (\tilde{x}_1, \tilde{x}_2)$$

Then, the corresponding fundamental matrix in the original coordinates is

$$F = T_2^\top \tilde{F} T_1$$

Proof. Substitute $\tilde{x}_1 = T_1 x_1$ and $\tilde{x}_2 = T_2 x_2$ into the normalised epipolar constraint:

$$(T_2 x_2)^\top \tilde{F} (T_1 x_1) = 0 \iff x_2^\top (T_2^\top \tilde{F} T_1) x_1 = 0$$

□

7.3 Rank-2 Enforcement

Proposition 7.5 (Closest rank-2 matrix in Frobenius norm). *Let $F = U\Sigma V^\top$ be the SVD with $\Sigma = \text{diag}(\sigma_1, \sigma_2, \sigma_3)$, $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq 0$. Define $\Sigma' = \text{diag}(\sigma_1, \sigma_2, 0)$ and $F' = U\Sigma'V^\top$. Then F' has rank 2 and minimises $\|F - G\|_F$ over all rank-2 matrices G .*

Proof. This is the Eckart–Young–Mirsky theorem: the best rank- k approximation of a matrix in Frobenius norm is obtained by truncating its SVD to the top k singular values. Take $k = 2$. □

8 Metric Reconstruction from the Essential Matrix

This section concerns the geometric reconstruction problem: given the essential matrix, recover the relative camera pose and reconstruct 3D scene points.

8.1 Decomposition of the Essential Matrix

Proposition 8.1 (Canonical SVD form). *Let $E = [t]_{\times} R$ be an essential matrix with $t \neq 0$. Then E has rank 2 and singular values $(\|t\|, \|t\|, 0)$. In particular, after scaling, one may write an SVD*

$$E = U\Sigma V^{\top}, \quad \Sigma = \text{diag}(1, 1, 0)$$

Proof. The rank and singular value structure were established earlier: $EE^{\top} = [t]_{\times}[t]_{\times}^{\top} = \|t\|^2 I - tt^{\top}$, which has eigenvalues $\|t\|^2, \|t\|^2, 0$. Hence the singular values of E are $\|t\|, \|t\|, 0$. Since E is defined up to nonzero scale, we may rescale so that the two nonzero singular values are 1, yielding $\Sigma = \text{diag}(1, 1, 0)$. \square

Theorem 8.2 (Pose recovery from E). *Let $E \in \mathbb{R}^{3 \times 3}$ be an essential matrix with SVD*

$$E = U\Sigma V^{\top}, \quad \Sigma = \text{diag}(1, 1, 0).$$

Define

$$W = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad Z = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Then, the relative rotation and translation direction satisfy

$$R \in \{UWV^{\top}, UW^{\top}V^{\top}\}, \quad [t]_{\times} = UZU^{\top}$$

Yielding four candidate poses $(R, \pm t)$, where t is determined only up to an unknown overall scale.

Proof. Since $\Sigma = \text{diag}(1, 1, 0)$, we may write

$$\Sigma = ZW^{\top}$$

Because a direct multiplication gives

$$ZW^{\top} = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \Sigma$$

Hence

$$E = U\Sigma V^{\top} = UZW^{\top}V^{\top}$$

Define $[t]_{\times} := UZU^{\top}$, which is skew-symmetric because Z is skew-symmetric and U is orthogonal.

Next define candidate rotations

$$R_1 := UWV^{\top}, \quad R_2 := UW^{\top}V^{\top}$$

Using orthogonality $U^{\top}U = V^{\top}V = I$, we compute

$$[t]_{\times}R_1 = (UZU^{\top})(UWV^{\top}) = UZWV^{\top}$$

Similarly,

$$[t]_{\times}R_2 = (UZU^{\top})(UW^{\top}V^{\top}) = UZW^{\top}V^{\top}$$

Since $E = UZW^{\top}V^{\top}$, we have $E = [t]_{\times}R_2$. The alternative choice R_1 corresponds to a second valid decomposition $E = [t]_{\times}R$ because changing W to W^{\top} produces the same essential manifold up to the intrinsic four-fold ambiguity.

Finally, $[t]_{\times}$ determines t only up to sign, since $[(-t)]_{\times} = -[t]_{\times}$, and E itself is defined only up to overall nonzero scale, so the magnitude of t cannot be recovered from two views alone. Combining the two candidate rotations with $\pm t$ yields four candidate poses. \square

8.2 Cheirality Condition

Proposition 8.3 (Cheirality to select the valid pose). *Let $P_1 = [I \mid 0]$ and $P_2 = [R \mid t]$ be candidate normalised camera matrices obtained from an essential matrix decomposition. For a set of correct correspondences in a non-degenerate configuration, exactly one of the four candidates $(R, \pm t)$ yields triangulated 3D points with positive depth in both camera frames (cheirality).*

Proof. For any candidate pose, triangulation yields a projective point $X \in \mathbb{P}^3$ which, when expressed in each camera frame, has a depth given by the third component of X_1 and $X_2 = RX_1 + t$ (up to homogeneous scale).

Physically valid reconstructions require that the point lies in front of both cameras, i.e. both depths are positive after choosing a consistent scale.

The four decompositions differ by a reflection of the reconstructed rays with respect to the baseline direction and/or reversal of the translation direction. In a non-degenerate configuration with multiple correspondences, only one choice places the majority of points in front of both cameras simultaneously. Hence cheirality uniquely selects the correct solution. \square

8.3 Triangulation

Proposition 8.4 (Linear triangulation system). *Given two projection matrices P_1, P_2 and corresponding homogeneous image points $x_1, x_2 \in \mathbb{P}^2$, the 3D point $X \in \mathbb{P}^3$ satisfies*

$$x_i \times (P_i X) = 0, \quad i = 1, 2$$

Each equation yields two independent linear constraints in X . Stacking both views produces a homogeneous system

$$AX = 0$$

Solvable (up to scale) by SVD as the right singular vector corresponding to the smallest singular value.

Proof. For fixed i , the relation $x_i \sim P_i X$ in \mathbb{P}^2 means that x_i and $P_i X$ are colinear, which is equivalent to $x_i \times (P_i X) = 0$. The cross product yields three linear equations in the homogeneous coordinates of X , but only two are independent because one equation is a linear combination of the other two due to homogeneity. Selecting two independent rows for each view and stacking them yields a matrix A such that $AX = 0$. Under noise, the least-squares solution is obtained as the right singular vector associated with the smallest singular value of A . \square

8.4 Error metrics for epipolar geometry

In practice, correspondences (x_1, x_2) are noisy, so the epipolar constraint, $x_2^\top F x_1 = 0$, will not hold exactly.

So, let $x_1, x_2 \in \mathbb{P}^2$ be corresponding homogeneous image points (typically with $w = 1$), and define the induced epipolar lines

$$\ell_2 := F x_1, \quad \ell_1 := F^\top x_2$$

Where ℓ_2 is the epipolar line in image 2 corresponding to x_1 , and ℓ_1 is the epipolar line in image 1 corresponding to x_2 .

8.4.1 Algebraic epipolar residual

Definition 8.5 (Algebraic residual). The algebraic epipolar residual of a correspondence (x_1, x_2) under F is

$$r(x_1, x_2; F) := x_2^\top F x_1$$

Proposition 8.6 (Scale sensitivity of the algebraic residual). *For any nonzero scalar α , one has*

$$r(x_1, x_2; \alpha F) = \alpha r(x_1, x_2; F)$$

Hence the algebraic residual is not invariant to the projective scaling ambiguity $F \sim \alpha F$

Proof. Immediate from bilinearity:

$$r(x_1, x_2; \alpha F) = x_2^\top (\alpha F) x_1 = \alpha x_2^\top F x_1$$

□

Scale-invariant residuals Because of this scale dependence, raw algebraic residuals are best used only after fixing a convention for the scale of F (e.g. $\|F\|_F = 1$), or as an intermediate quantity inside scale-invariant metrics.

8.4.2 Point-to-line distance and symmetric epipolar distance

Lemma 8.7 (Euclidean point-to-line distance in homogeneous coordinates). *Let $\ell = (a, b, c)^\top \in \mathbb{P}^2$ represent the line*

$$\ell = \{(u, v) \in \mathbb{R}^2 : au + bv + c = 0\}$$

Let $x = (u, v, 1)^\top$ be a point in homogeneous coordinates. If $(a, b) \neq (0, 0)$, then the Euclidean distance from x to ℓ is

$$d(x, \ell) = \frac{|\ell^\top x|}{\sqrt{a^2 + b^2}}$$

Proof. Let $n := (a, b)^\top \in \mathbb{R}^2$, and write the line as

$$n^\top y + c = 0, \quad y = (y_1, y_2)^\top \in \mathbb{R}^2$$

Since $(a, b) \neq (0, 0)$, we have $n \neq 0$. Define the unit normal vector

$$\hat{n} := \frac{n}{\|n\|}, \quad \|n\| = \sqrt{a^2 + b^2}$$

We claim that for any point $p = (u, v)^\top$, the (signed) distance from p to the line is

$$d_{\text{signed}}(p, \ell) = \hat{n}^\top p + \frac{c}{\|n\|}$$

To prove this, consider the orthogonal projection of p onto the line along the normal direction. Any point on the normal line through p has the form

$$q(\lambda) = p - \lambda \hat{n}, \quad \lambda \in \mathbb{R}$$

We choose λ such that $q(\lambda) \in \ell$, i.e.

$$n^\top q(\lambda) + c = 0$$

Substitute $q(\lambda) = p - \lambda \hat{n}$:

$$n^\top(p - \lambda \hat{n}) + c = 0 \iff n^\top p - \lambda n^\top \hat{n} + c = 0$$

But $n^\top \hat{n} = n^\top(n/\|n\|) = \|n\|$

Hence

$$n^\top p - \lambda \|n\| + c = 0 \implies \lambda = \frac{n^\top p + c}{\|n\|}$$

Now observe that λ is exactly the signed length of the displacement from p to its projection $q(\lambda)$, because \hat{n} is a unit vector:

$$\|p - q(\lambda)\| = \|\lambda \hat{n}\| = |\lambda|$$

Therefore the Euclidean distance is

$$d(p, \ell) = |\lambda| = \frac{|n^\top p + c|}{\|n\|} = \frac{|au + bv + c|}{\sqrt{a^2 + b^2}}$$

Finally, writing $x = (u, v, 1)^\top$ and $\ell = (a, b, c)^\top$ gives $au + bv + c = \ell^\top x$, so

$$d(x, \ell) = \frac{|\ell^\top x|}{\sqrt{a^2 + b^2}}$$

□

Definition 8.8 (Symmetric epipolar distance). Given a correspondence (x_1, x_2) and a fundamental matrix F , the induced epipolar lines are

$$\ell_2 := Fx_1 \quad (\text{the epipolar line in image 2}), \quad \ell_1 := F^\top x_2 \quad (\text{the epipolar line in image 1})$$

The *symmetric epipolar distance* measures how well the correspondence satisfies the epipolar geometry by summing the squared point-to-line distances in both images:

$$d_{\text{sym}}^2(x_1, x_2; F) := d(x_2, \ell_2)^2 + d(x_1, \ell_1)^2$$

In other words, it penalises both deviations: how far x_2 lies from the line predicted by x_1 , and how far x_1 lies from the line predicted by x_2 .

If $x_i = (u_i, v_i, 1)^\top$, $\ell_2 = (a_2, b_2, c_2)^\top$, and $\ell_1 = (a_1, b_1, c_1)^\top$, then using the Euclidean point-to-line distance formula one may write

$$d_{\text{sym}}^2(x_1, x_2; F) = \frac{(x_2^\top Fx_1)^2}{a_2^2 + b_2^2} + \frac{(x_1^\top Fx_2)^2}{a_1^2 + b_1^2}$$

Proposition 8.9 (Scale invariance of symmetric epipolar distance). *For any nonzero scalar α , the symmetric epipolar distance satisfies*

$$d_{\text{sym}}^2(x_1, x_2; \alpha F) = d_{\text{sym}}^2(x_1, x_2; F)$$

Proof. Scaling $F \mapsto \alpha F$ scales the epipolar lines $\ell_2 \mapsto \alpha \ell_2$, $\ell_1 \mapsto \alpha \ell_1$.

Then the numerators $(\ell^\top x)^2$ scale by α^2 , and the denominators $a^2 + b^2$ also scale by α^2 . Hence, each squared distance term is invariant, and so is their sum. □

8.4.3 Sampson distance (first-order geometric approximation)

The symmetric epipolar distance treats the two point-to-line distances separately, but it is still not the true geometric reprojection error. A widely used approximation is the Sampson distance, obtained by a first-order (Gauss–Newton) correction of the epipolar constraint.

Definition 8.10 (Epipolar constraint function). Let $x_1 = (u_1, v_1, 1)^\top$, $x_2 = (u_2, v_2, 1)^\top$.

Define the scalar constraint

$$g(u_1, v_1, u_2, v_2) := x_2^\top F x_1$$

Theorem 8.11 (Sampson distance formula). *Let g be as above and assume $\nabla g \neq 0$ at the correspondence.*

The Sampson distance is defined by

$$d_{\text{samp}}^2(x_1, x_2; F) := \frac{g^2}{\|\nabla g\|^2}$$

Where the gradient is taken with respect to (u_1, v_1, u_2, v_2) .

Writing $\ell_2 = F x_1 = (a_2, b_2, c_2)^\top$ and $\ell_1 = F^\top x_2 = (a_1, b_1, c_1)^\top$, one obtains the explicit expression

$$d_{\text{samp}}^2(x_1, x_2; F) = \frac{(x_2^\top F x_1)^2}{a_2^2 + b_2^2 + a_1^2 + b_1^2}$$

Proof. Consider the smallest correction $\delta = (\delta u_1, \delta v_1, \delta u_2, \delta v_2)^\top$ such that the corrected correspondence satisfies the epipolar constraint to first order.

Linearising g around the observation gives

$$g(u + \delta u) \approx g(u) + \nabla g(u)^\top \delta$$

We seek the minimum-norm correction δ such that the linearised constraint is satisfied:

$$g(u) + \nabla g(u)^\top \delta = 0, \quad \text{minimise } \|\delta\|^2$$

This is a constrained least-squares problem. By standard Lagrange multiplier arguments, the minimiser is attained when δ is parallel to ∇g , i.e. $\delta = -\lambda \nabla g$. Substituting into the constraint yields

$$g(u) - \lambda \|\nabla g(u)\|^2 = 0 \quad \implies \quad \lambda = \frac{g(u)}{\|\nabla g(u)\|^2}$$

Therefore

$$\|\delta\|^2 = \lambda^2 \|\nabla g(u)\|^2 = \frac{g(u)^2}{\|\nabla g(u)\|^2}$$

This defines the Sampson distance.

It remains to compute ∇g . Writing $g = x_2^\top F x_1$, differentiation with respect to the inhomogeneous coordinates gives

$$\frac{\partial g}{\partial u_2} = (F x_1)_1 = a_2, \quad \frac{\partial g}{\partial v_2} = (F x_1)_2 = b_2$$

and similarly

$$\frac{\partial g}{\partial u_1} = (F^\top x_2)_1 = a_1, \quad \frac{\partial g}{\partial v_1} = (F^\top x_2)_2 = b_1$$

Hence $\|\nabla g\|^2 = a_2^2 + b_2^2 + a_1^2 + b_1^2$, yielding the stated closed form. \square

Proposition 8.12 (Scale invariance of Sampson distance). *For any nonzero scalar α ,*

$$d_{\text{samp}}^2(x_1, x_2; \alpha F) = d_{\text{samp}}^2(x_1, x_2; F).$$

Proof. Under $F \mapsto \alpha F$, the numerator $(x_2^\top F x_1)^2$ scales by α^2 . The line coefficients $\ell_2 = F x_1$ and $\ell_1 = F^\top x_2$ scale by α , so the denominator $a_2^2 + b_2^2 + a_1^2 + b_1^2$ scales by α^2 . The ratio is invariant. \square

Aggregating per-correspondence errors Given correspondences $\{(x_1^{(i)}, x_2^{(i)})\}_{i=1}^N$ and a chosen metric $d^2(\cdot)$, a common summary statistic is the root-mean-square (RMS) error

$$\text{RMS}(d) = \sqrt{\frac{1}{N} \sum_{i=1}^N d^2(x_1^{(i)}, x_2^{(i)}; F)}$$

8.5 RANSAC for robust fundamental matrix estimation

In real data, putative correspondences $\{(x_1^{(i)}, x_2^{(i)})\}_{i=1}^N$ contain outliers (mismatches), so a single least-squares fit (e.g. the normalised 8-point algorithm) can be corrupted.

A standard remedy is RANSAC (Random Sample Consensus), which repeatedly fits a model from minimal random subsets and selects the model that attains the largest consensus set under an error threshold.

8.5.1 Minimal solver and inlier test

Let s denote the sample size required by a minimal solver. For the 8-point algorithm, $s = 8$. Given a candidate fundamental matrix F , define a per-correspondence error $d(\cdot)$ (e.g. Sampson distance or symmetric epipolar distance).

Declare i an inlier if for a chosen threshold $\tau > 0$

$$d(x_1^{(i)}, x_2^{(i)}; F) \leq \tau$$

The consensus set (inlier set) is

$$\mathcal{I}(F) := \{i \in \{1, \dots, N\} : d(x_1^{(i)}, x_2^{(i)}; F) \leq \tau\}$$

Definition 8.13 (RANSAC estimator for F). Fix an integer $T \geq 1$ (number of trials), a sample size s , and a threshold τ . For each trial $t = 1, \dots, T$:

- (i) Sample a subset $S_t \subset \{1, \dots, N\}$ uniformly at random with $|S_t| = s$.
- (ii) Fit F_t using a minimal solver on $\{(x_1^{(i)}, x_2^{(i)}) : i \in S_t\}$.
- (iii) Compute the consensus set $\mathcal{I}(F_t)$ under the chosen metric d and threshold τ .

Return the candidate with maximal consensus size:

$$F_{\text{best}} \in \arg \max_{t \in \{1, \dots, T\}} |\mathcal{I}(F_t)|$$

Optionally, refit by recomputing F from all the inliers $\mathcal{I}(F_{\text{best}})$.

Practical tie-breaking If multiple candidates achieve the same inlier count, a common tie-break is the smallest mean inlier error:

$$\frac{1}{|\mathcal{I}(F_t)|} \sum_{i \in \mathcal{I}(F_t)} d(x_1^{(i)}, x_2^{(i)}; F_t)$$

8.5.2 Success probability and iteration bound

RANSAC succeeds once it draws at least one all-inlier minimal sample.

Definition 8.14 (Inlier ratio). Let $w \in [0, 1]$ denote the (unknown) fraction of inliers among the N correspondences:

$$w := \frac{|\mathcal{I}_\star|}{N}$$

Where \mathcal{I}_\star is the true inlier set (with respect to an ideal model and noise level).

Lemma 8.15 (Probability of an all-inlier sample). *Assume each RANSAC trial samples s indices uniformly at random, and that the inlier ratio is w . Then the probability that a single trial draws an all-inlier sample is*

$$p_{\text{good}} = w^s$$

Proof. Under the idealised independence model, the probability that one randomly drawn correspondence is an inlier is w . For an s -tuple of draws, the probability all s are inliers is the product w^s . Note: When sampling without replacement, the exact probability differs slightly but is well-approximated by w^s when N is large. \square

Theorem 8.16 (RANSAC iteration bound for a desired confidence). *Fix a desired success probability $p \in (0, 1)$. If the inlier ratio is w and the minimal sample size is s , then choosing the number of trials T to satisfy*

$$T \geq \frac{\log(1 - p)}{\log(1 - w^s)}$$

This ensures that the probability of drawing at least one all-inlier sample is at least p .

Proof. Let $p_{\text{good}} = w^s$ be the probability that a single trial is all-inlier. The probability a trial is not all-inlier is $1 - p_{\text{good}}$.

Assuming independent trials, the probability that none of the T trials is all-inlier is

$$(1 - p_{\text{good}})^T = (1 - w^s)^T$$

Therefore, the probability of at least one all-inlier sample is

$$1 - (1 - w^s)^T$$

To achieve success probability at least p , require

$$1 - (1 - w^s)^T \geq p \iff (1 - w^s)^T \leq 1 - p$$

Taking logarithms (note $0 < 1 - w^s < 1$, hence $\log(1 - w^s) < 0$) yields

$$T \log(1 - w^s) \leq \log(1 - p) \iff T \geq \frac{\log(1 - p)}{\log(1 - w^s)}$$

\square

Adaptive iteration count Since w is unknown, RANSAC commonly updates an empirical estimate $\hat{w} = |\mathcal{I}(F_{\text{best}})|/N$ as better models are found, and recomputes a target T via the bound above to stop early.

8.5.3 Why refit on inliers improves the estimate

The best model found by RANSAC is usually fitted from a minimal subset, which is statistically inefficient under noise. After identifying inliers, one should refit using all inliers.

Proposition 8.17 (Refitting on inliers reduces variance (heuristic)). *Assume the inlier correspondences are generated from a fixed ground-truth fundamental matrix F_\star with i.i.d. measurement noise, and assume the deterministic estimator \hat{F} behaves like a least-squares estimator on the chosen coordinates. Then, refitting F on a larger set of inliers typically reduces the estimator variance relative to using a single minimal sample.*

Proof. In least-squares estimation under i.i.d. noise, the covariance of the parameter estimate is proportional to $(A^\top A)^{-1}$, where A is the design matrix formed from the data. Using more (independent) inlier correspondences increases $A^\top A$ in the positive semidefinite order, which decreases $(A^\top A)^{-1}$ and hence reduces estimator variance. While the fundamental matrix estimation problem includes a rank constraint and projective scale, the same principle applies: fitting on more inliers yields a more stable estimate than fitting on a minimal subset. \square

9 Homography estimation and degeneracy testing

While the fundamental matrix captures general two-view geometry, a homography is the appropriate model for planar scenes and for (near) pure rotations.

9.1 Homography model

Definition 9.1 (Homography). A homography is a projective mapping $H \in \mathbb{R}^{3 \times 3}$, defined up to nonzero scale, that relates corresponding homogeneous image points $\tilde{x}_1, \tilde{x}_2 \in \mathbb{P}^2$ via

$$\tilde{x}_2 \sim H\tilde{x}_1$$

The notation $H \sim \alpha H$ for $\alpha \neq 0$ means that multiplying H by a nonzero scalar produces the same projective map, since homogeneous points are defined only up to scale.

Proposition 9.2 (Degrees of freedom and minimal sample size). *A homography H has 8 degrees of freedom (since $H \sim \alpha H$ for $\alpha \neq 0$). Each correspondence $\tilde{x}_2 \sim H\tilde{x}_1$ provides two independent constraints, so a minimal solver requires 4 point correspondences $(\tilde{x}_1^{(i)}, \tilde{x}_2^{(i)})$ in general position (not all colinear in either image).*

Proof. A 3×3 matrix has 9 entries, but the projective scaling ambiguity reduces this to 8 independent parameters. For one correspondence, $\tilde{x}_2 \sim H\tilde{x}_1$ is equivalent to $\tilde{x}_2 \times (H\tilde{x}_1) = 0$, which yields three linear equations but only two are independent due to homogeneity. Hence 4 correspondences yield 8 independent linear constraints, sufficient to determine H up to scale in generic configurations. \square

9.2 Planar scene check

Proposition 9.3 (Homography explains planar scenes and pure rotations). *If all observed 3D points lie on a plane $\pi : n^\top X + d = 0$ in the camera-1/world frame, then the image correspondence satisfies*

$$\tilde{x}_2 \sim H\tilde{x}_1, \quad H = K_2 \left(R - \frac{tn^\top}{d} \right) K_1^{-1}$$

In the special case of pure rotation ($t = 0$), the correspondence is also a homography:

$$\tilde{x}_2 \sim K_2 R K_1^{-1} \tilde{x}_1$$

Proof. For a 3D point X in the camera-1/world frame, the calibrated projection equations are

$$\lambda_1 \tilde{x}_1 = K_1 X, \quad \lambda_2 \tilde{x}_2 = K_2 (RX + t)$$

Assume X lies on the plane π , so $n^\top X = -d$. Then,

$$\left(R - \frac{tn^\top}{d} \right) X = RX - t \frac{n^\top X}{d} = RX - t \frac{-d}{d} = RX + t$$

Substituting into the camera-2 projection gives

$$\lambda_2 \tilde{x}_2 \sim K_2 \left(R - \frac{tn^\top}{d} \right) X$$

From camera 1, $X \sim K_1^{-1} \tilde{x}_1$, so

$$\tilde{x}_2 \sim K_2 \left(R - \frac{tn^\top}{d} \right) K_1^{-1} \tilde{x}_1 =: H \tilde{x}_1$$

Setting $t = 0$ yields $\tilde{x}_2 \sim K_2 R K_1^{-1} \tilde{x}_1$, the pure-rotation case. □

9.3 Homography estimation via Direct Linear Transform

Proposition 9.4 (Linear constraints for homography (DLT)). *Let $\tilde{x}_1 = (u_1, v_1, 1)^\top$ and $\tilde{x}_2 = (u_2, v_2, 1)^\top$ be corresponding homogeneous image points. Writing the rows of H as $h_1^\top, h_2^\top, h_3^\top$, the relation $\tilde{x}_2 \sim H\tilde{x}_1$ yields two independent linear equations in the entries of H . Stacking $N \geq 4$ correspondences gives a homogeneous system*

$$Ah = 0, \quad h := \text{vec}(H) \in \mathbb{R}^9,$$

solvable (up to scale) by SVD.

Proof. The projective relation $\tilde{x}_2 \sim H\tilde{x}_1$ means there exists a nonzero scalar λ such that

$$\lambda \tilde{x}_2 = H \tilde{x}_1$$

Equivalently, \tilde{x}_2 and $H\tilde{x}_1$ are colinear, which is expressed by the cross product constraint

$$\tilde{x}_2 \times (H\tilde{x}_1) = 0$$

Rather than expanding the cross product directly, it is convenient to write

$$H\tilde{x}_1 = \begin{bmatrix} h_1^\top \tilde{x}_1 \\ h_2^\top \tilde{x}_1 \\ h_3^\top \tilde{x}_1 \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

Since $\tilde{x}_2 = (u_2, v_2, 1)^\top$ and $\tilde{x}_2 \sim (a, b, c)^\top$, dehomogenisation implies

$$u_2 = \frac{a}{c}, \quad v_2 = \frac{b}{c}$$

Which is equivalent to the two equations

$$a - u_2 c = 0, \quad b - v_2 c = 0$$

Substituting $a = h_1^\top \tilde{x}_1$, $b = h_2^\top \tilde{x}_1$, $c = h_3^\top \tilde{x}_1$ yields two equations

$$h_1^\top \tilde{x}_1 - u_2 h_3^\top \tilde{x}_1 = 0, \quad h_2^\top \tilde{x}_1 - v_2 h_3^\top \tilde{x}_1 = 0$$

These are linear in the unknown rows h_1, h_2, h_3 . Writing them in block-matrix form gives

$$\begin{bmatrix} \tilde{x}_1^\top & 0 & -u_2 \tilde{x}_1^\top \\ 0 & \tilde{x}_1^\top & -v_2 \tilde{x}_1^\top \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix} = Ah = 0$$

For each correspondence i , we obtain two such rows; stacking all N correspondences produces

$$Ah = 0, A \in \mathbb{R}^{2N \times 9}, \quad h = \text{vec}(H) \in \mathbb{R}^9$$

With noisy measurements, an exact null vector typically does not exist, so one solves the constrained least-squares problem

$$\min_{\|h\|_2=1} \|Ah\|_2$$

Let $A = U\Sigma V^\top$ be the SVD with singular values $\sigma_1 \geq \dots \geq \sigma_9 \geq 0$. Writing $y = V^\top h$ and using orthogonality of U, V , we have

$$\|Ah\|_2 = \|U\Sigma V^\top h\|_2 = \|\Sigma y\|_2$$

Thus

$$\|Ah\|_2^2 = \sum_{k=1}^9 \sigma_k^2 y_k^2, \quad \text{with } \|y\|_2 = \|h\|_2 = 1$$

This is minimised by placing all mass on the smallest singular value, i.e. $y = e_9$, so the minimiser is the right singular vector corresponding to the smallest singular value:

$$h = V e_9$$

Reshaping h into a 3×3 matrix yields H , defined up to scale. □

Hartley Normalisation As with the normalised 8-point algorithm, DLT is numerically improved by normalising image coordinates via similarity transforms T_1, T_2 so that the centroid is near the origin and the mean distance to the origin is $\sqrt{2}$ - also known as Hartley normalisation. One estimates \hat{H} in normalised coordinates and denormalises via $H = T_2^{-1} \hat{H} T_1$.

9.3.1 Homography error metrics

Definition 9.5 (Reprojection and transfer error). Given $\tilde{x}_2 \sim H\tilde{x}_1$, define the predicted point in image 2 by

$$\hat{\tilde{x}}_2 := H\tilde{x}_1, \quad \hat{x}_2 := \pi(\hat{\tilde{x}}_2)$$

Where $\pi(a, b, c)^\top := (a/c, b/c)^\top$ is dehomogenisation.

The (forward) reprojection/transfer error is

$$e_{1 \rightarrow 2}(x_1, x_2; H) := \|x_2 - \hat{x}_2\|_2$$

Similarly define $e_{2 \rightarrow 1}(x_2, x_1; H^{-1})$.

The symmetric transfer error is

$$e_{\text{sym}}^2(x_1, x_2; H) := e_{1 \rightarrow 2}(x_1, x_2; H)^2 + e_{2 \rightarrow 1}(x_2, x_1; H^{-1})^2$$

Proposition 9.6 (Scale invariance of the symmetric transfer error). *For any nonzero scalar α , one has*

$$e_{\text{sym}}^2(x_1, x_2; \alpha H) = e_{\text{sym}}^2(x_1, x_2; H)$$

Proof. Scaling $H \mapsto \alpha H$ scales the homogeneous prediction $\hat{\tilde{x}}_2 = H\tilde{x}_1$ by α , but dehomogenisation π is invariant to overall scale: $\pi(\alpha\hat{\tilde{x}}_2) = \pi(\hat{\tilde{x}}_2)$. Hence both forward and backward reprojection terms are unchanged, and so is their sum. \square

9.4 RANSAC for robust homography estimation

Homography estimation from real correspondences is corrupted by outliers, so we apply RANSAC with a minimal sample size $s = 4$.

Definition 9.7 (RANSAC estimator for H). Fix a number of trials T , a sample size $s = 4$, and a threshold $\tau_H > 0$. For each trial $t = 1, \dots, T$:

- (i) Sample $S_t \subset \{1, \dots, N\}$ uniformly at random with $|S_t| = 4$
- (ii) Fit H_t using a DLT solver on $\{(x_1^{(i)}, x_2^{(i)}) : i \in S_t\}$
- (iii) Define the inlier set by a transfer-error test, e.g.

$$\mathcal{I}(H_t) := \left\{ i : e_{\text{sym}}^2(x_1^{(i)}, x_2^{(i)}; H_t) \leq \tau_H^2 \right\}$$

Return $H_{\text{best}} \in \arg \max_t |\mathcal{I}(H_t)|$, optionally refitting H on all inliers.

9.5 Planar / pure-rotation degeneracy detection for two-view bootstrap

Definition 9.8 (Homography-vs-fundamental model selection). Run RANSAC for both a fundamental matrix F (or essential matrix E) and a homography H , producing inlier sets \mathcal{I}_F and \mathcal{I}_H on the same correspondence set.

Declare a candidate pair homography-dominant if for a margin $\gamma > 1$

$$|\mathcal{I}_H| \geq \gamma |\mathcal{I}_F|$$

Proposition 9.9 (Interpretation of homography dominance). *If $|\mathcal{I}_H|$ substantially exceeds $|\mathcal{I}_F|$, then the correspondences are better explained by a planar projective mapping than by general epipolar geometry. This is strong evidence of an (approximately) planar scene, (approximately) pure rotation, or very distant structure, and two-view monocular initialisation is likely ill-conditioned.*

Proof. In planar scenes and pure-rotation motion, the correspondence relation is exactly (or approximately) $\tilde{x}_2 \sim H\tilde{x}_1$, so a homography achieves a large consensus set under a geometric error threshold. In contrast, epipolar geometry relies on a non-degenerate baseline and non-planar structure to constrain depth; when parallax is small, fitting F can yield fewer inliers and the recovered pose/triangulation becomes unstable. Hence homography dominance is a practical diagnostic for rejecting initialisation pairs where triangulation is numerically unstable and many reconstructed points are pushed towards infinity. \square