Robust Methods of Computing Fundamental Matrices

Presented by Phil David

Papers:

Z. Zhang. Determining the Epipolar Geometry and its Uncertainty: A Review. *International Journal of Computer Vision*, 27(2), pp. 161-195, 1998. Also INRIA Research Report No. 2927.

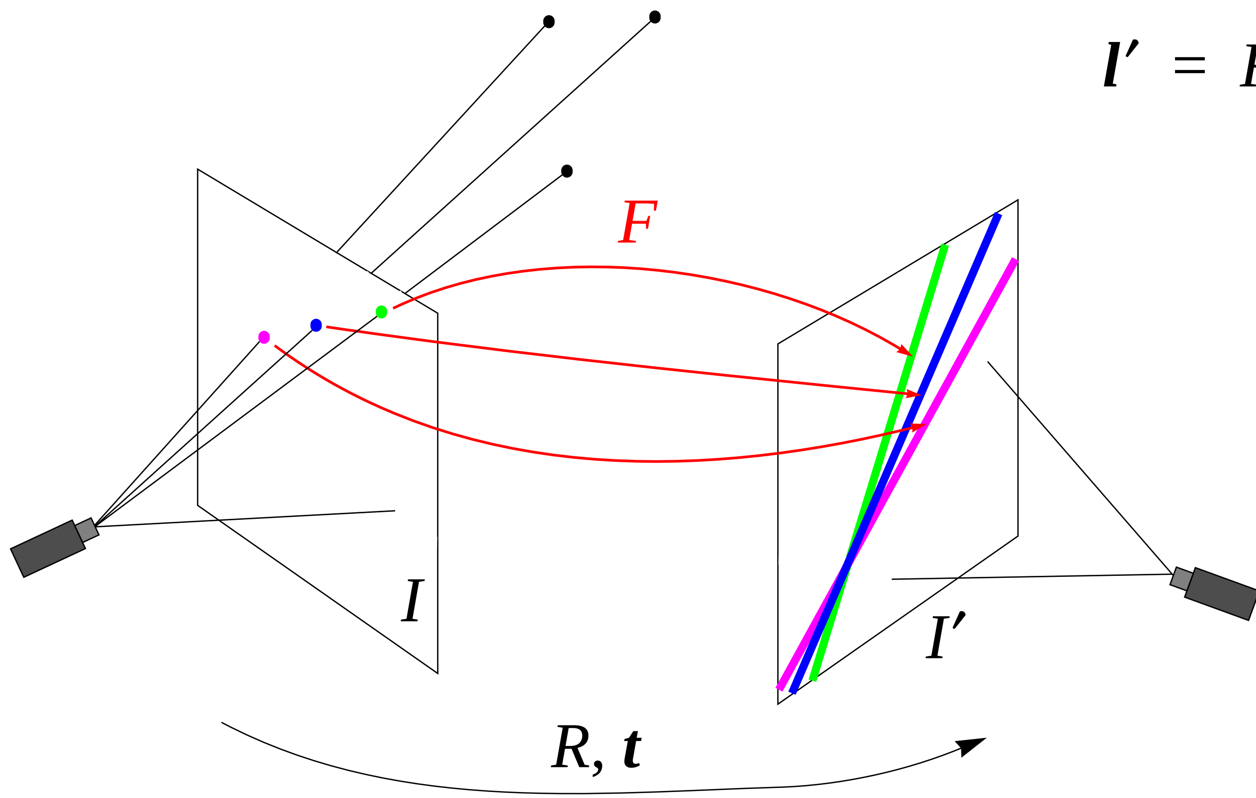
O. Faugeras and S. Laveau. 3D Scene Representation as a Collection of Images and Fundamental Matrices. *Proceedings of ICPR94*, pp. 689-691, 1994. Also INRIA Research Report No. 2205.

Outline

1. Background material: projective geometry.
2. Derivation of the fundamental matrix.
3. Computing the fundamental matrix from point correspondences.
4. Uncertainty in the fundamental matrix.
5. 3D scene representation using fundamental matrices.

The Fundamental Matrix

The fundamental matrix fully specifies the constraints on the possible motions of corresponding image points between two views, when no other information about the scene is known.



$$\boldsymbol{l}' = F\boldsymbol{x}$$

Why Compute a Fundamental Matrix?

It has many uses in multi-view computer vision:

- Narrows the search for corresponding points.
- Used in algorithms for:
  - ✦ Camera calibration,
  - ✦ Motion determination and segmentation,
  - ✦ 3D reconstruction of scenes.
  - ✦ Generation of new views of a scene.

Homogeneous Coordinates

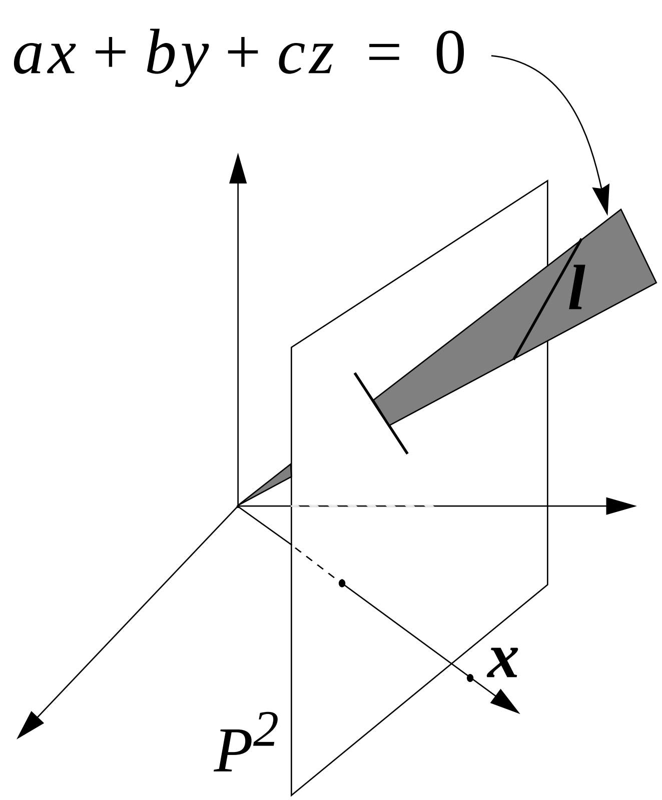
An n-dimensional inhomogeneous (Euclidean) point is represented with n+1 coordinates.

Scene points: $$\boldsymbol{X} = \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \Rightarrow \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = \begin{pmatrix} WX \\ WY \\ WZ \\ W \end{pmatrix} \text{ for } W \neq 0.$$

Image points: $$\boldsymbol{x} = \begin{pmatrix} x \\ y \end{pmatrix} \Rightarrow \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} wx \\ wy \\ w \end{pmatrix} \text{ for } w \neq 0.$$

Planar Projective Geometry

In projective geometry, all objects (e.g., points, lines, planes, etc.) that have the same image are considered equivalent.



Points:

$$\mathbf{x} = (x, y)^\top \Rightarrow (x, y, 1)^\top = (wx, wy, w)^\top \text{ for } w \neq 0.$$

Lines:

$$ax + by + c = 0$$

$$(a, b, c)^\top (x, y, 1) = 0$$

$$(sa, sb, sc)^\top (wx, wy, w) = 0$$

$$\mathbf{l} = (sa, sb, sc)^\top \text{ for } s \neq 0$$

$$\mathbf{l}^\top \mathbf{x} = 0$$

Planar Projective Geometry

A point $\boldsymbol{x}$ lies on a line $\boldsymbol{l}$ iff $\boldsymbol{x}^\top \boldsymbol{l} = \boldsymbol{l}^\top \boldsymbol{x} = 0$.

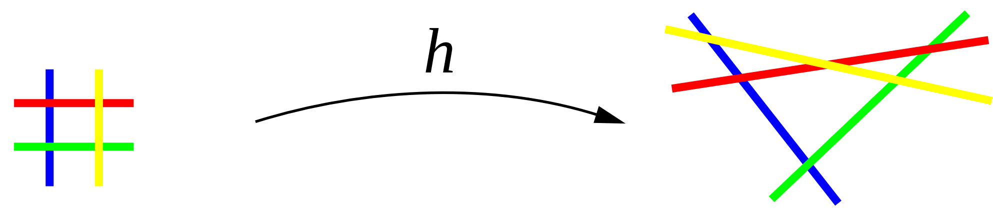
The intersection of two lines $\boldsymbol{l}$ and $\boldsymbol{l}'$ is the point $\boldsymbol{x} = \boldsymbol{l} \times \boldsymbol{l}'$.

Proof: $\boldsymbol{x}^\top \boldsymbol{l} = 0$ and $\boldsymbol{x}^\top \boldsymbol{l}' = 0$, so the vector $\boldsymbol{x}$ is perpendicular to both $\boldsymbol{l}$ and $\boldsymbol{l}'$, so $\boldsymbol{x} = \boldsymbol{l} \times \boldsymbol{l}'$.

The line through two points $\boldsymbol{x}$ and $\boldsymbol{x}'$ is $\boldsymbol{l} = \boldsymbol{x} \times \boldsymbol{x}'$.

Planar Projective Transformations

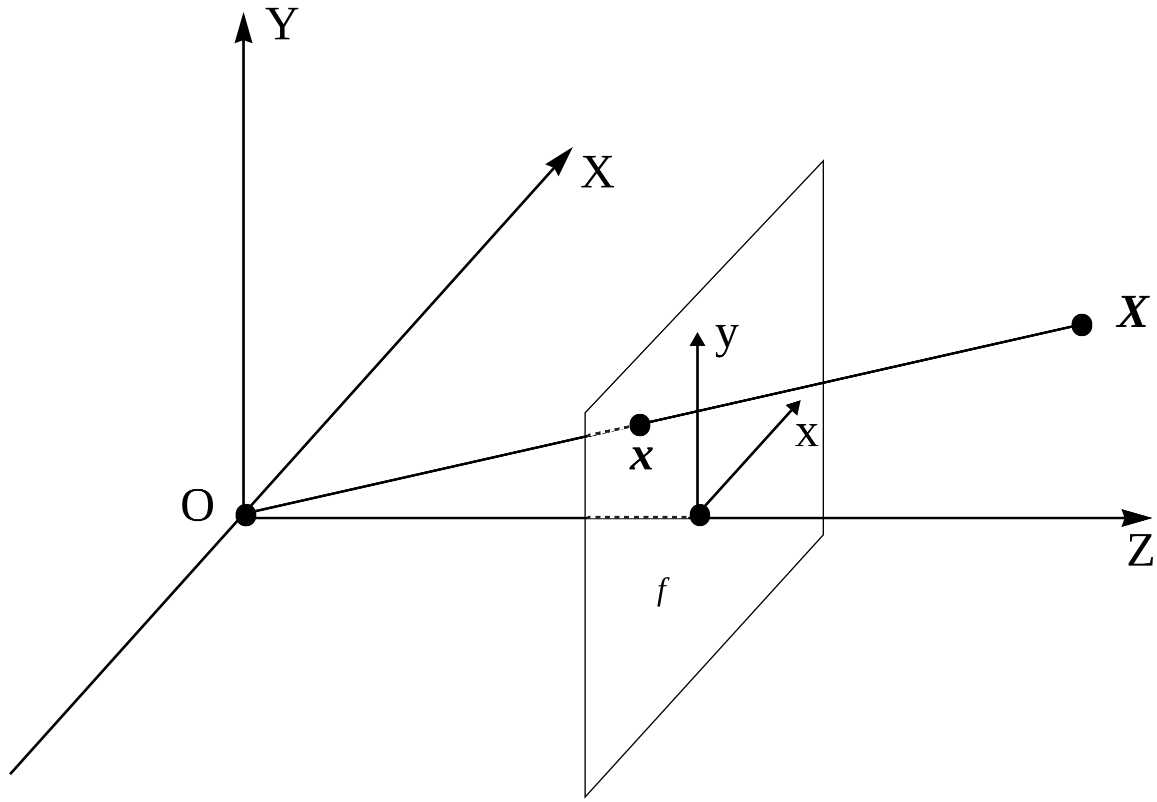
A **planar projective transformation** $h: P^2 \rightarrow P^2$ is a invertible mapping such that three points $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, and $\boldsymbol{x}_3$ lie on the same line if and only if $h(\boldsymbol{x}_1)$, $h(\boldsymbol{x}_2)$, and $h(\boldsymbol{x}_3)$ lie on the same line.



A mapping $h: P^2 \rightarrow P^2$ is a planar projective transformation if and only if there exists a nonsingular 3×3 matrix H such that for any point $\boldsymbol{x} \in P^2$, $h(\boldsymbol{x}) = H(\boldsymbol{x})$.

Projective transformations are also known as **homographies**, **projectivities**, and **collineations**.

Pinhole Camera Model



$$x = \frac{fX}{Z}$$

$$y = \frac{fY}{Z}$$

Perspective Projection Equations

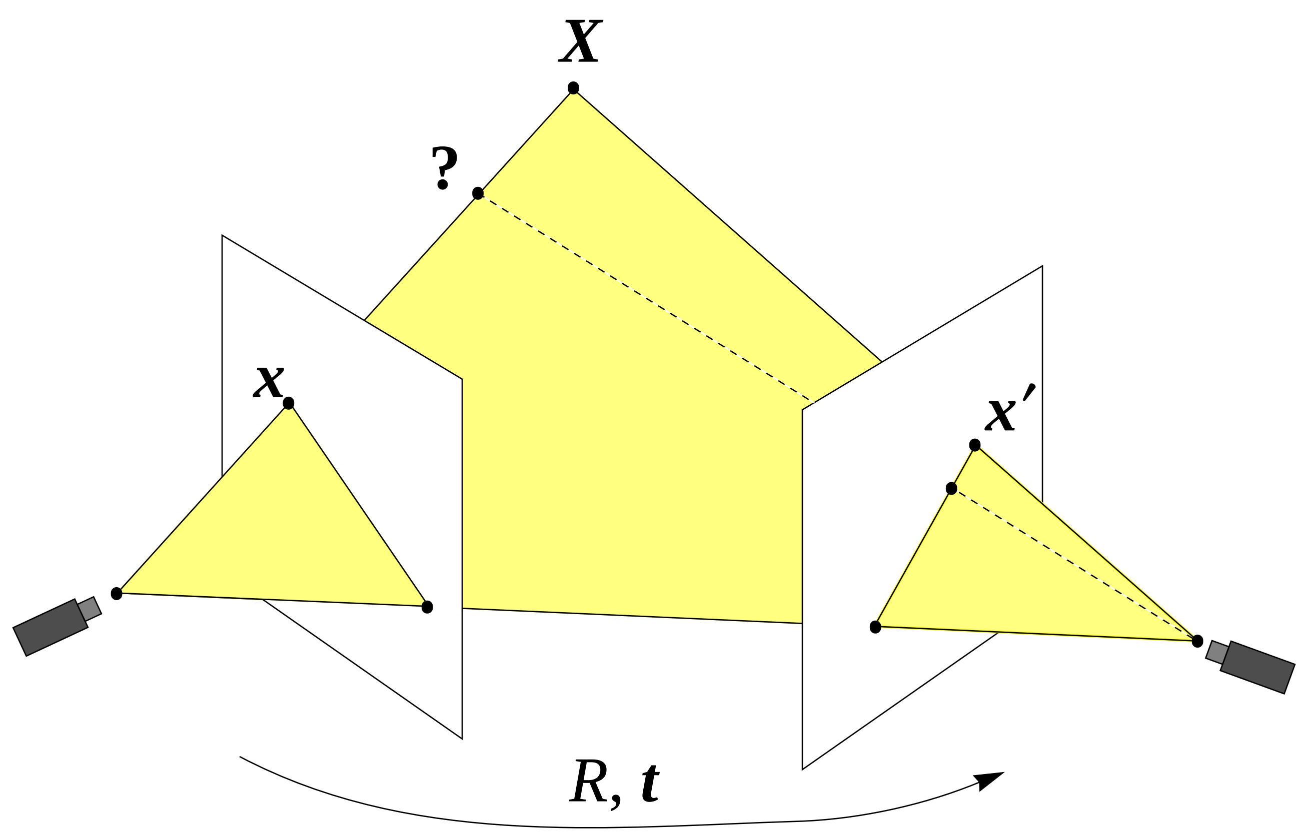
$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{bmatrix} -fk_x & 0 & x_0 \\ 0 & -fk_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & \boldsymbol{t} \\ \mathbf{0}^\top & 1 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

$$= K[R \ \ \boldsymbol{t}]\boldsymbol{X}.$$

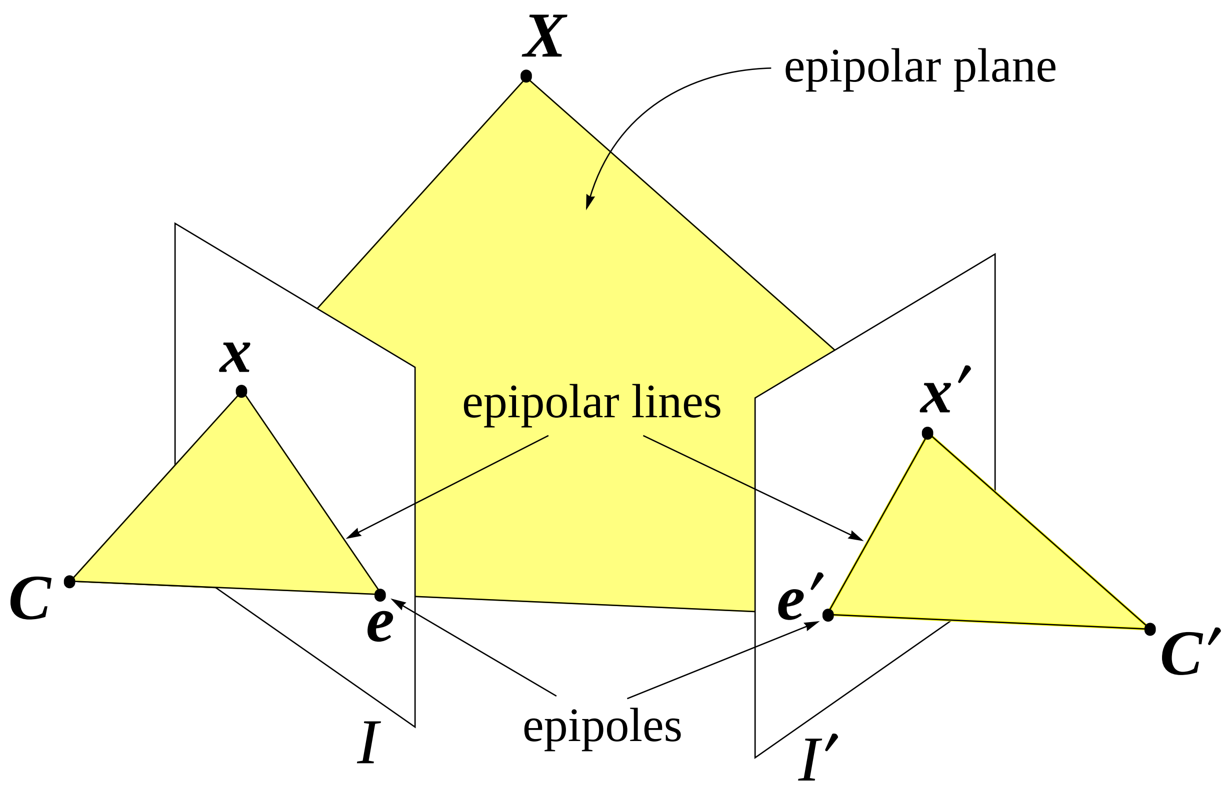
$$\begin{aligned} \boldsymbol{X}_{\text{camera}} &= T_{\text{Euclidean}} \boldsymbol{X}_{\text{world}} \\ \boldsymbol{x}_{\text{image}} &= [I \ \ \mathbf{0}] \boldsymbol{X}_{\text{camera}} \\ \boldsymbol{x}_{\text{pixel}} &= K \boldsymbol{x}_{\text{image}} \end{aligned}$$

Epipolar Geometry

Two perspective images of a rigid scene are related by what is called epipolar geometry.

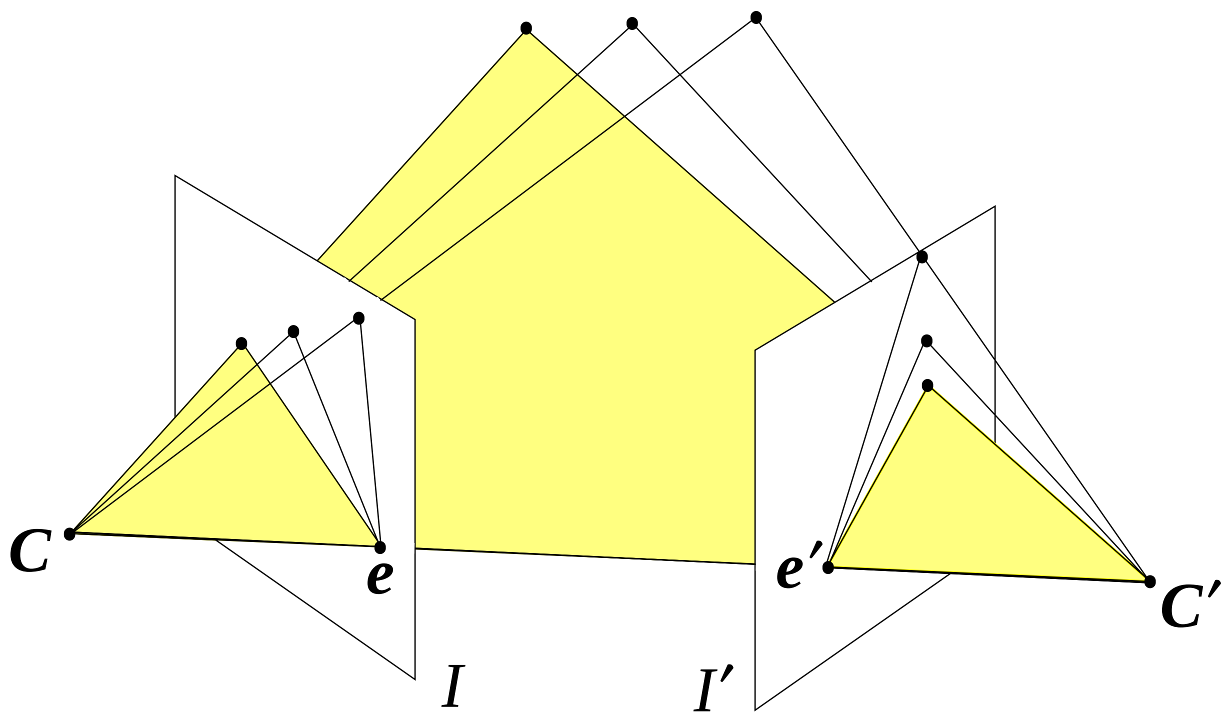


Epipolar Geometry

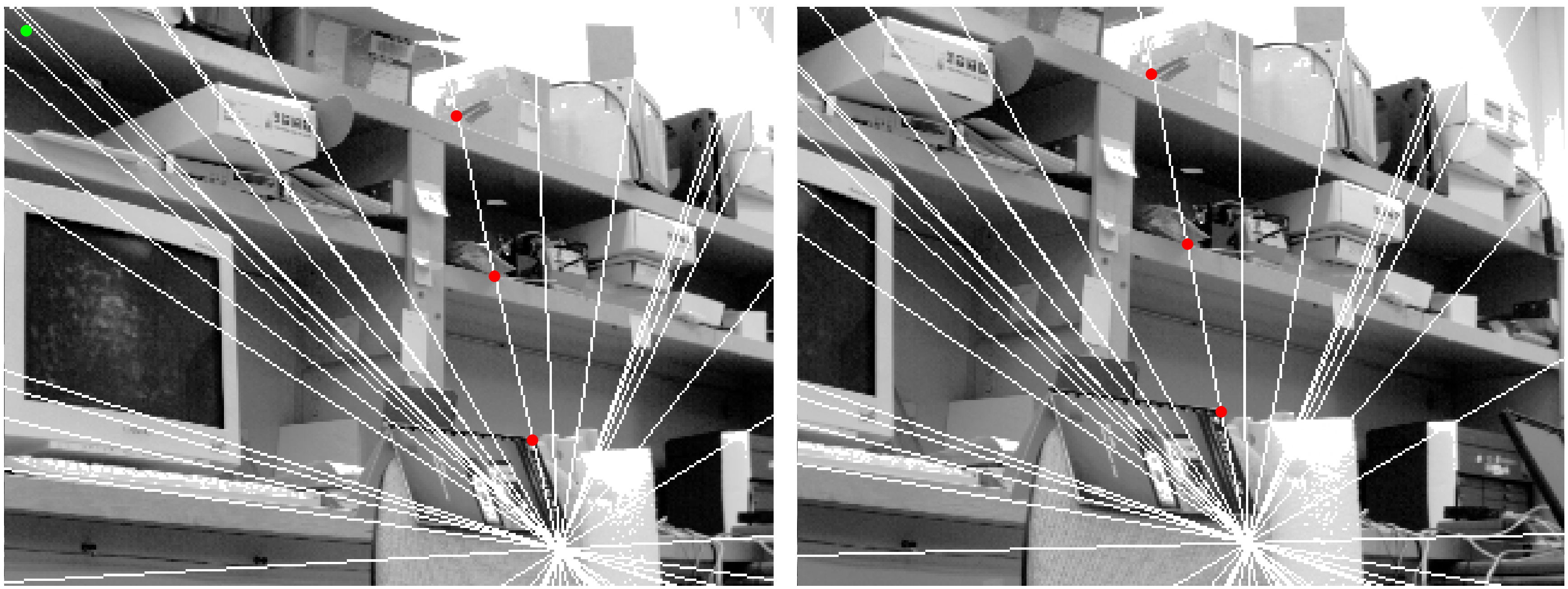


Epipolar constraint

Corresponding points must lie on corresponding epipolar lines.



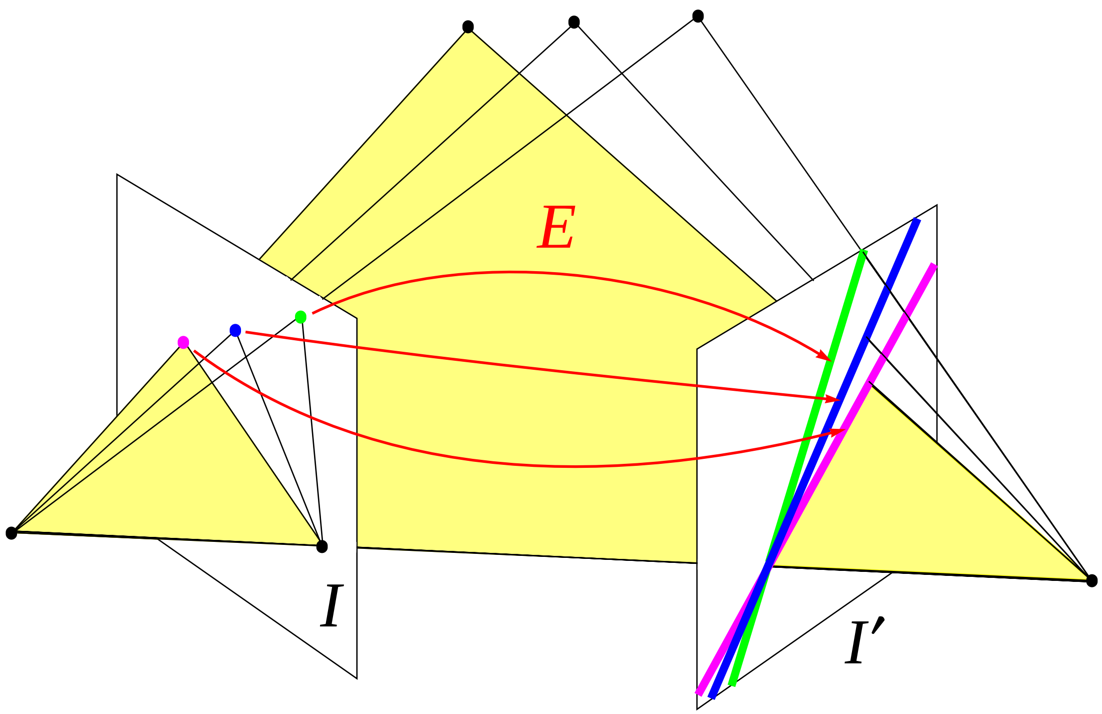
Example Image Pair



The Essential Matrix

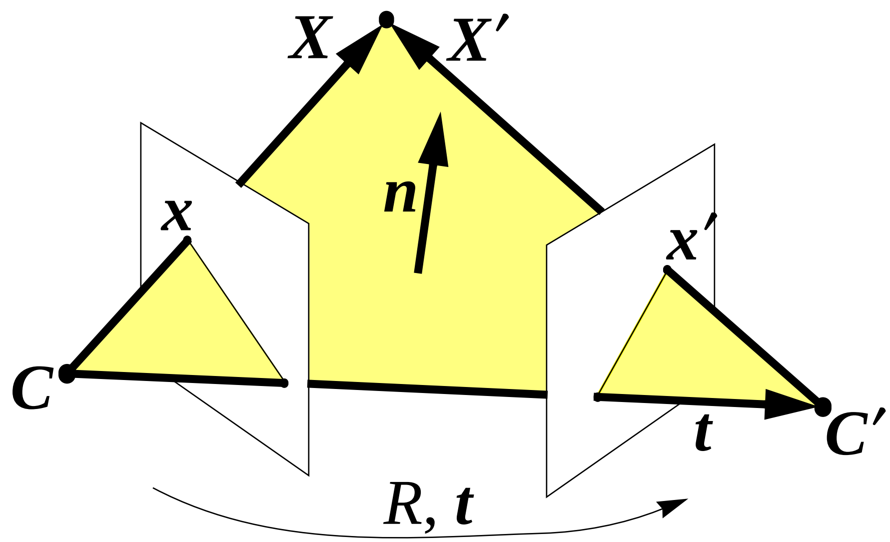
- Maps points in one image into the corresponding epipolar lines in the other image.

- Points are given in terms of camera coordinates, not pixel coordinates.



$$\boldsymbol{l}' = E\boldsymbol{x}$$

$\boldsymbol{X}'$, $\boldsymbol{t}$, and $R\boldsymbol{X}$ are coplanar...



$$\boldsymbol{X}' = R\boldsymbol{X} + \boldsymbol{t}$$

(all 3-vectors)

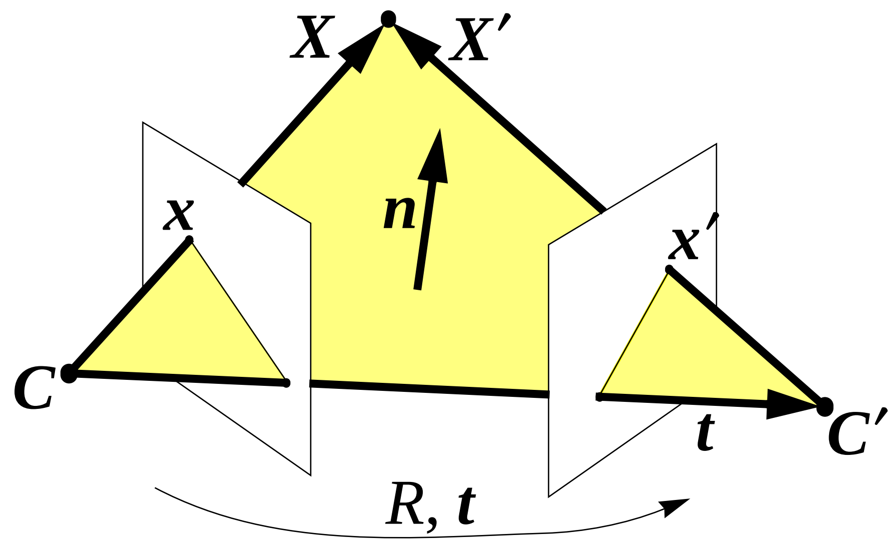
If $\boldsymbol{n}$ is the normal to plane spanned by $\boldsymbol{X}'$ and $\boldsymbol{t}$, then

$$\begin{aligned} \boldsymbol{n}^\top \boldsymbol{X}' &= 0 \\ &= \boldsymbol{n}^\top (R\boldsymbol{X} + \boldsymbol{t}) \\ &= \boldsymbol{n}^\top R\boldsymbol{X} + \boldsymbol{n}^\top \boldsymbol{t} \\ &= \boldsymbol{n}^\top R\boldsymbol{X} \end{aligned}$$

So, $\boldsymbol{X}'$, $\boldsymbol{t}$, and $R\boldsymbol{X}$ are coplanar.

The Essential Matrix

$\boldsymbol{X}' = R\boldsymbol{X} + \boldsymbol{t}$



$\boldsymbol{X}'$, $\boldsymbol{t}$, and $R\boldsymbol{X}$ are coplanar.

Normal to epipolar plane is:

$$\boldsymbol{n} = \boldsymbol{t} \times (R\boldsymbol{X}).$$

Then, $\boldsymbol{X}'^{\top}\boldsymbol{n} = 0 = \boldsymbol{X}'^{\top}\boldsymbol{t} \times (R\boldsymbol{X}) = \boldsymbol{X}'^{\top}E\boldsymbol{X}$ where

$$E = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix} R, \quad \text{since} \quad \boldsymbol{t} \times \boldsymbol{v} = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix} \boldsymbol{v} \quad \text{for any } \boldsymbol{v}.$$

The Essential Matrix

$$\boldsymbol{X'}^{\top} E \boldsymbol{X} = 0 \quad \Rightarrow \quad (X', Y', Z') E \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = 0$$

$$x = \frac{fX}{Z} \qquad y = \frac{fY}{Z} \qquad x' = \frac{f'X'}{Z'} \qquad y' = \frac{f'X'}{Z'}$$

Multiple by $\frac{f'f}{Z'Z}$, you get $(f'X'/Z', f'Y'/Z', f')E \begin{pmatrix} \frac{fX}{Z} \\ \frac{fY}{Z} \\ f \end{pmatrix} = 0$

or, $(x', y', f')E \begin{pmatrix} x \\ y \\ f \end{pmatrix} = 0$.

The Epipolar Constraint

$$(x', y', f')E\begin{pmatrix} x \\ y \\ f \end{pmatrix} = 0 \text{ or } \boldsymbol{x}'^{\top}E\boldsymbol{x} = 0$$

Let:

$\boldsymbol{x}$ be a point in image I,

$\boldsymbol{x}'$ be the corresponding point in image I', and

$\boldsymbol{l}' = E\boldsymbol{x}$

Then:

$\boldsymbol{x}'^{\top}\boldsymbol{l}' = 0$

$\boldsymbol{l}'$ gives the homogeneous coordinates of the line in I' on which the corresponding point $\boldsymbol{x}'$ must lie, in terms of image coordinates.

The Fundamental Matrix

Recall $\boldsymbol{x}_{\text{pixel}} = K\boldsymbol{x}_{\text{image}}$

So,

$$\tilde{\boldsymbol{x}} = K\boldsymbol{x} \text{ and } \tilde{\boldsymbol{x}}' = K'\boldsymbol{x}'$$

$$\boldsymbol{x} = K^{-1}\tilde{\boldsymbol{x}} \text{ and } \boldsymbol{x}' = K'^{-1}\tilde{\boldsymbol{x}}'.$$

Substituting into $\boldsymbol{x}'^{\mathsf{T}}E\boldsymbol{x} = 0$ gives

$$\tilde{\boldsymbol{x}}'^{\mathsf{T}}K'^{-\mathsf{T}}EK^{-1}\tilde{\boldsymbol{x}} = 0.$$

Let $F = K'^{-\mathsf{T}}EK^{-1}$, then

$$\tilde{\boldsymbol{x}}'^{\mathsf{T}}F\tilde{\boldsymbol{x}} = 0.$$

The Epipolar Constraint

$$\tilde{\mathbf{x}}'^{\top} F \tilde{\mathbf{x}} = 0$$

Let:

$\tilde{\mathbf{x}}$ be a point in image I,

$\tilde{\mathbf{x}}'$ be the corresponding point in image I', and

$\mathbf{l}' = F \tilde{\mathbf{x}}$

Then:

$\tilde{\mathbf{x}}'^{\top} \mathbf{l}' = 0$

$\mathbf{l}'$ gives the homogeneous coordinates of the line in I' on which the corresponding point $\tilde{\mathbf{x}}'$ must lie, in terms of pixel coordinates.

Properties of the Fundamental Matrix

1. F is defined up to a scale factor: if $\tilde{\boldsymbol{x}}'^{\top} F \tilde{\boldsymbol{x}} = 0$, then $\tilde{\boldsymbol{x}}'^{\top} (\alpha F) \tilde{\boldsymbol{x}} = 0$ for any $\alpha \neq 0$.

2. F has rank 2.

3. F has 7 degrees of freedom: a 3×3 homogeneous matrix has 8 degrees of freedom since it is defined only up to a scale factor; but, since F satisfies $\det(F) = 0$, one degree of freedom is lost.

4. If $F : I \rightarrow I'$ then $F^{\top} : I' \rightarrow I$.

5. When the internal parameters of two cameras are known, the fundamental matrix for the pair of images is transformed in to the essential matrix for that pair of cameras.

Degenerate Configurations

Scene points on a quadric surface: When all of the scene points lie on a quadric surface passing through the two centers of projection, then there may be three different fundamental matrices compatible with the data. The two sets of image points are more specifically related by a quadratic transformation:

$$\tilde{\boldsymbol{x}}' = F_1 \tilde{\boldsymbol{x}} \times F_2 \tilde{\boldsymbol{x}}$$

where F_1 and F_2 are two of the fundamental matrices.

Rotating camera: In the case when the camera rotates about its center of projection and does not translate, the most constraining epipolar geometry that can be determined for corresponding image points is given by a homography:

$$\tilde{\boldsymbol{x}}' = H \tilde{\boldsymbol{x}}.$$

Calculating the Fundamental Matrix

Given a set of n corresponding pixels in two images:

$$\begin{pmatrix} u_i \\ v_i \end{pmatrix} \leftrightarrow \begin{pmatrix} u_i' \\ v_i' \end{pmatrix}, \quad i = 1, \ldots, n\,.$$

These correspond to the homogeneous points:

$$\begin{pmatrix} u_i \\ v_i \\ 1 \end{pmatrix} \leftrightarrow \begin{pmatrix} u_i' \\ v_i' \\ 1 \end{pmatrix}, \quad i = 1, \ldots, n\,.$$

Let $\boldsymbol{u}_i = (u_i, v_i, 1)^\top$ and $\boldsymbol{u}_i' = (u_i', v_i', 1)^\top$.

Find the 3×3 matrix F such that $\boldsymbol{u}_i'^\top F \boldsymbol{u}_i = 0$ for $i = 1, \ldots, n$.

Methods of Calculating the F Matrix

1. Linear methods

2. Nonlinear methods

3. Robust methods

    Random Sample Consensus - RANSAC

    Least Median of Squares - LMS

Linear Method - The 8-Point Algorithm

$\boldsymbol{u}_i'^{\top} F \boldsymbol{u}_i = 0$ is linear in the elements of F:

$$u_i u_i' F_{11} + v_i u_i' F_{12} + u_i' F_{13} + u_i v_i' F_{21} + v_i v_i' F_{22} + v_i' F_{23} + u_i F_{31} + v_i F_{32} + F_{33} = 0.$$

This can be written as

$$(u_i u_i', v_i u_i', u_i', u_i v_i', v_i v_i', v_i', u_i, v_i, 1)^{\top} \boldsymbol{f} = 0$$

where

$$\boldsymbol{f} = (F_{11}, F_{12}, F_{13}, F_{21}, F_{22}, F_{23}, F_{31}, F_{32}, F_{33})^{\top}.$$

The 8-Point Algorithm

Given n point correspondences, $\boldsymbol{u}_i \leftrightarrow \boldsymbol{u}_i'$, $i = 1, \ldots, n$, the n linear equations may be written as

$$A\boldsymbol{f} = \boldsymbol{0}$$

where A is an $n \times 9$ matrix and the i-th row of A is

$$(u_i u_i', v_i u_i', u_i', u_i v_i', v_i v_i', v_i', u_i, v_i, 1).$$

Since $\boldsymbol{f} = \boldsymbol{0}$ is of no use, require that $\|\boldsymbol{f}\| = 1$.

Since $\boldsymbol{f}$ has 8 degrees of freedom ($\det(F) = 0$ not enforced), we need at least 8 point correspondences.

There will be a unique nonzero solution provided A has rank 8.

Noise in the Data

When there is noise in the measured locations of corresponding points, the matrix A will have rank 9. In this case, there will not be a nonzero solution to $A\boldsymbol{f} = \mathbf{0}$.

Instead, minimize $\|A\boldsymbol{f}\|^2$ subject to $\|\boldsymbol{f}\|^2 = 1$.

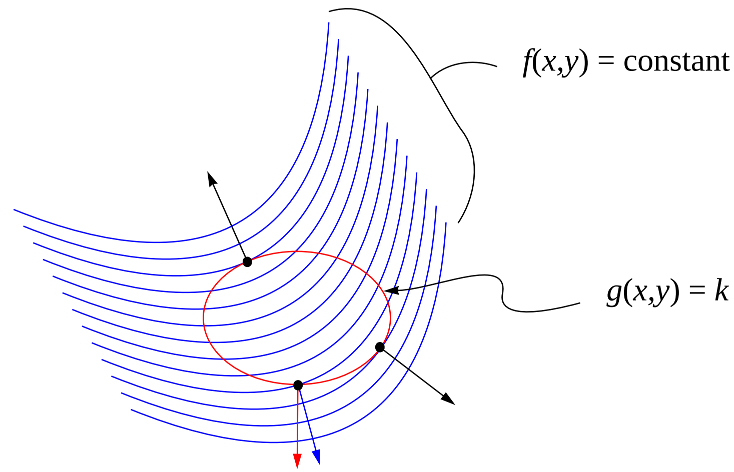
$$\left(\min_F \sum_i (\boldsymbol{u}_i'^\top F \boldsymbol{u}_i)^2 \text{ subject to } \|F\|^2 = 1\right)$$

The solution to this problem is the unit eigenvector corresponding to the smallest eigenvalue of $A^\top A$.

This is easily computed by singular value decomposition (SVD).

Lagrange Multipliers

Optimize $f(x, y)$ subject to $g(x, y) = k$:



Necessary conditions for a solution at $(\hat{x}, \hat{y})$:

$$\nabla f(\hat{x}, \hat{y}) \text{ is parallel to } \nabla g(\hat{x}, \hat{y}) \text{ and } g(\hat{x}, \hat{y}) = k$$

$$\nabla f(\hat{x}, \hat{y}) = \lambda \nabla g(\hat{x}, \hat{y}) \text{ and } g(\hat{x}, \hat{y}) = k$$

$$\nabla f(\hat{x}, \hat{y}) - \lambda \nabla g(\hat{x}, \hat{y}) = 0 \text{ and } g(\hat{x}, \hat{y}) = k$$

Minimize $\|A\boldsymbol{f}\|^2$ subject to $\|\boldsymbol{f}\|^2 = 1$

Solve: $\frac{d}{d\boldsymbol{f}}\|A\boldsymbol{f}\|^2 - \lambda\frac{d}{d\boldsymbol{f}}\|\boldsymbol{f}\| = 0$ and $\|\boldsymbol{f}\|^2 = 1$

Substitute $\|A\boldsymbol{f}\|^2 = (A\boldsymbol{f})^\mathsf{T}A\boldsymbol{f} = \boldsymbol{f}^\mathsf{T}A^\mathsf{T}A\boldsymbol{f}$ and $\|\boldsymbol{f}\|^2 = \boldsymbol{f}^\mathsf{T}\boldsymbol{f}$

$L(\boldsymbol{f}, \lambda) = \boldsymbol{f}^\mathsf{T}A^\mathsf{T}A\boldsymbol{f} + \lambda\boldsymbol{f}^\mathsf{T}\boldsymbol{f}$ $\qquad$ $\frac{d}{d\boldsymbol{f}}L(\boldsymbol{f}, \lambda) = A^\mathsf{T}A\boldsymbol{f} - \lambda\boldsymbol{f}$

We seek $\hat{\boldsymbol{f}}$ and $\hat{\lambda}$ such that $A^\mathsf{T}A\hat{\boldsymbol{f}} - \hat{\lambda}\hat{\boldsymbol{f}} = 0$ or $A^\mathsf{T}A\hat{\boldsymbol{f}} = \hat{\lambda}\hat{\boldsymbol{f}}$.
$\hat{\boldsymbol{f}}$ is an eigenvector of $A^\mathsf{T}A$ corresponding to eigenvalue $\hat{\lambda}$.

$\|A\hat{f}\|^2 = \hat{\lambda}$ (using $\boldsymbol{f}^\mathsf{T}\boldsymbol{f} = 1$), so choose the smallest eigenvalue and corresponding eigenvector.

Rank 2 Constraint

The constraint that $\mathrm{rank}(F) = 2$ was not enforced.

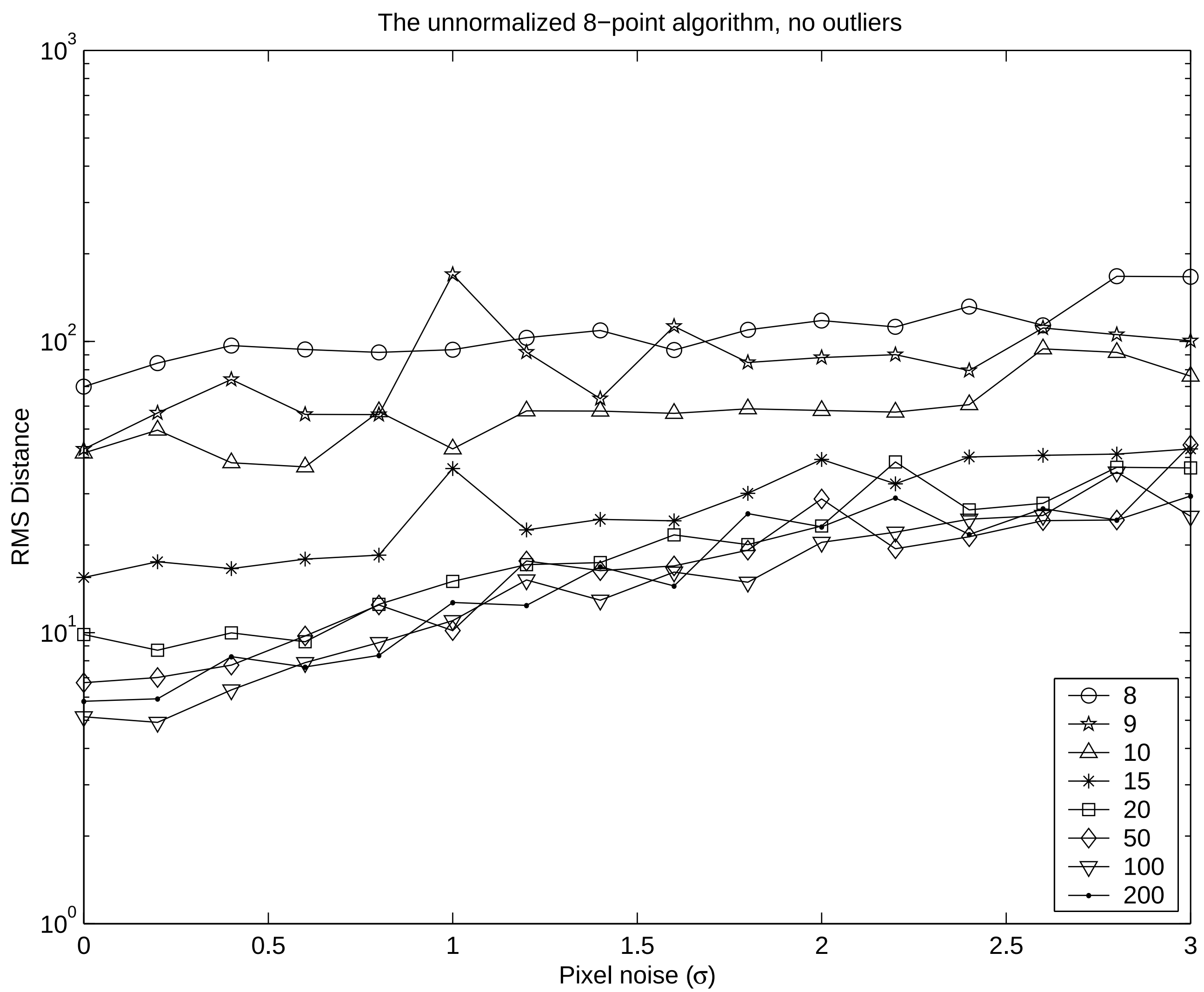
Replace F with F' such that $\|F - F'\|_{\mathrm{F}}$ is minimized and $\det(F') = 0$:

1. $F = UDV^{\mathsf{T}}$
   (singular value decomposition, $D = \mathrm{diag}(\sigma_1, \sigma_2, \sigma_3)$)

2. $F' = U \cdot \mathrm{diag}(\sigma_1, \sigma_2, 0) \cdot V^{\mathsf{T}}$

The Frobenius norm: $\|A\|_{\mathrm{F}} = \sum_{i,j} a_{ij}^2$

Performance of the 8-Point Algorithm



Problem: Ill-Conditioned

The problem of minimizing $\|A\boldsymbol{f}\|^2$ subject to $\|\boldsymbol{f}\|^2 = 1$ is ill-conditioned.

Ill-conditioned: small changes to the data (input) can produce large changes in the answers.

This linear method of computing F is ill-conditioned when the coordinates of corresponding pixels span a few orders of magnitude (e.g., $0 - 512$ in value). To improve the conditioning of the problem, the coordinates of pixels are normalized so that, on average, points have coordinates near $(1, 1, 1)^\top$ (i.e., are a distance of $\sqrt{2}$ from the origin). [Hartley, 1997]

Normalization of the Data

1. Transform the homogeneous image coordinates, each image separately, according to $\tilde{\boldsymbol{u}}_i = T\boldsymbol{u}_i$ and $\tilde{\boldsymbol{u}}_i' = T'\boldsymbol{u}_i'$ such that the new points lie, on average, a distance of $\sqrt{2}$ from the origin. For a set of inhomogeneous points $\{\boldsymbol{x}_i\}$ in the first image, the matrix T that accomplished this normalization is

$$T = \begin{bmatrix} \frac{\sqrt{2}}{d} & 0 & -\frac{\sqrt{2}c_x}{d} \\ 0 & \frac{\sqrt{2}}{d} & -\frac{\sqrt{2}c_y}{d} \\ 0 & 0 & 1 \end{bmatrix}$$

where $(c_x, c_y)^\mathsf{T}$ is the centroid of the points $\{\boldsymbol{x}_i\}$, and d is the average distance of these points from this centroid.

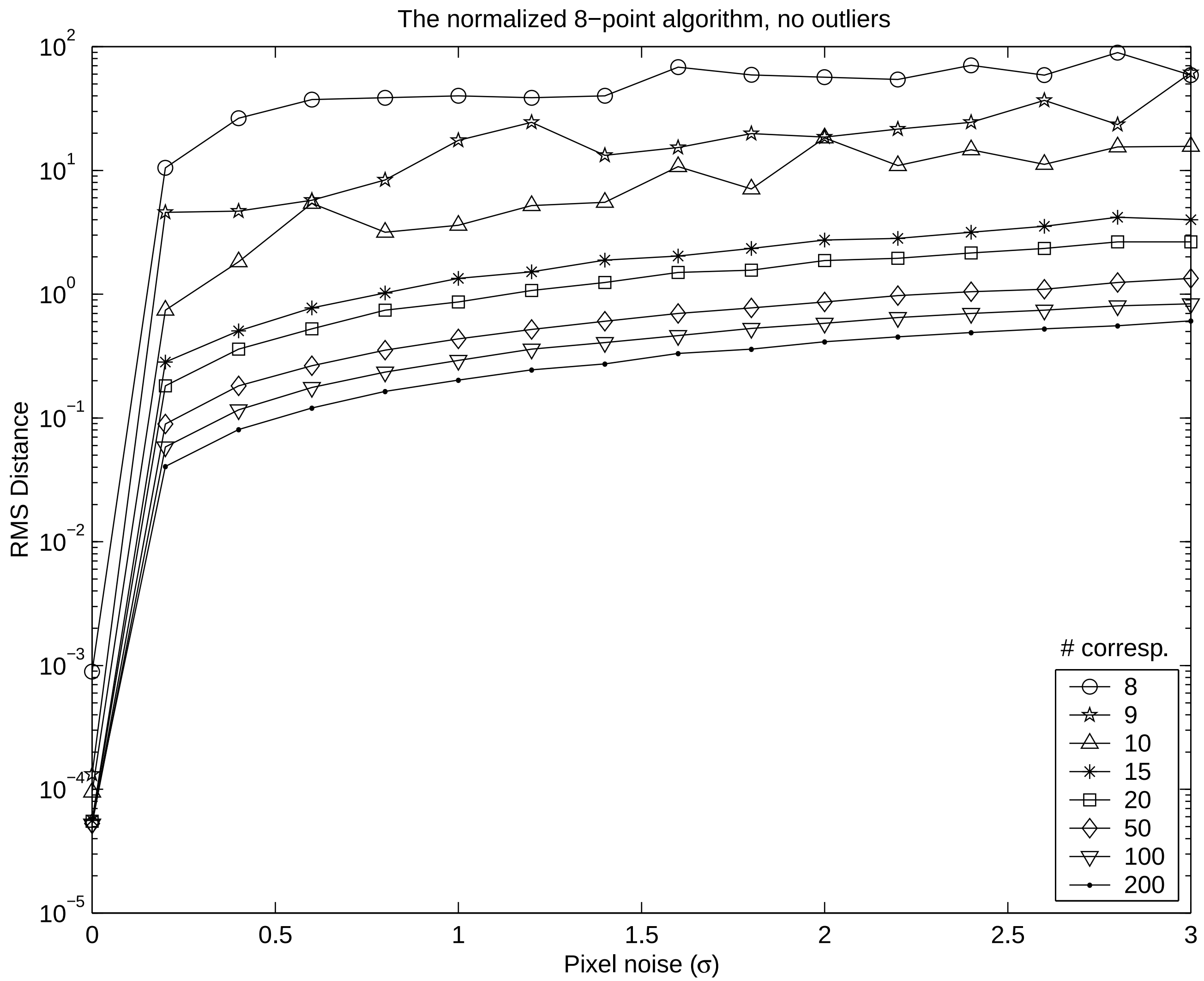
Normalization of the Data...

2. The matrix T' that normalizes the points $\{\boldsymbol{x}_i'\}$ in the second image is calculated similarly.

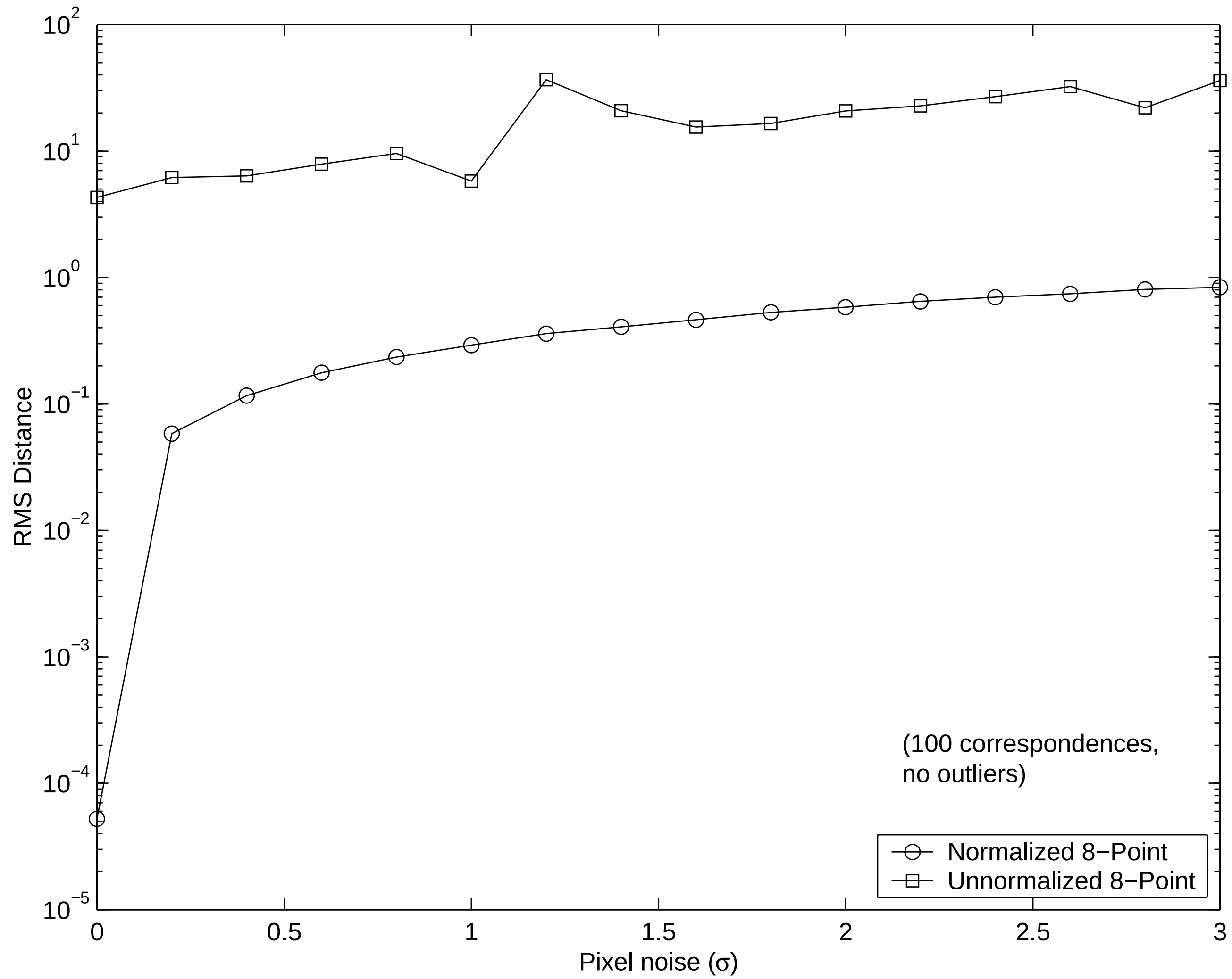
3. Find the fundamental matrix $\tilde{F}$ associated with the correspondences $\{\tilde{\boldsymbol{u}}_i \leftrightarrow \tilde{\boldsymbol{u}}_i'\}$.

4. The fundamental matrix associated with the original set of correspondences is $F = T'^{\top}\tilde{F}T$.

Performance of the Normalized 8-Pnt. Alg.



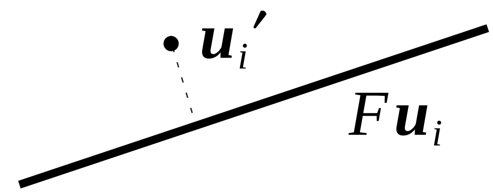
Comparison



10^{2}
10^{1}
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
RMS Distance
0
0.5
1
1.5
2
2.5
3
Pixel noise (σ)
(100 correspondences,
no outliers)
Normalized 8−Point
Unnormalized 8−Point

Algebraic vs Geometric Error Functions

Given a point correspondence: $\boldsymbol{u}_i \leftrightarrow \boldsymbol{u}_i'$



The Euclidean distance of the point $\boldsymbol{u}_i' = (u_i, v_i, 1)^\mathsf{T}$ from its computed epipolar line $\boldsymbol{l}_i' = (l_{i1}', l_{i2}', l_{i3}')^\mathsf{T} = F\boldsymbol{u}_i$ is

$$d(\boldsymbol{u}_i', \boldsymbol{l}_i') = \frac{\boldsymbol{u}_i'^\mathsf{T}\boldsymbol{l}_i'}{\sqrt{l_{i1}'^2 + l_{i2}'^2}} = \frac{\boldsymbol{u}_i'^\mathsf{T}F\boldsymbol{u}_i}{\sqrt{l_{i1}'^2 + l_{i2}'^2}}$$

Algebraic error: $\min_F \sum_i (\boldsymbol{u}_i'^\mathsf{T}F\boldsymbol{u}_i)^2 = (l_{i1}'^2 + l_{i2}'^2)d^2(\boldsymbol{u}_i', \boldsymbol{l}_i')$

Geometric error: $d^2(\boldsymbol{u}_i', \boldsymbol{l}_i')$

Minimization of Geometric Error

Use a nonlinear minimization algorithm.

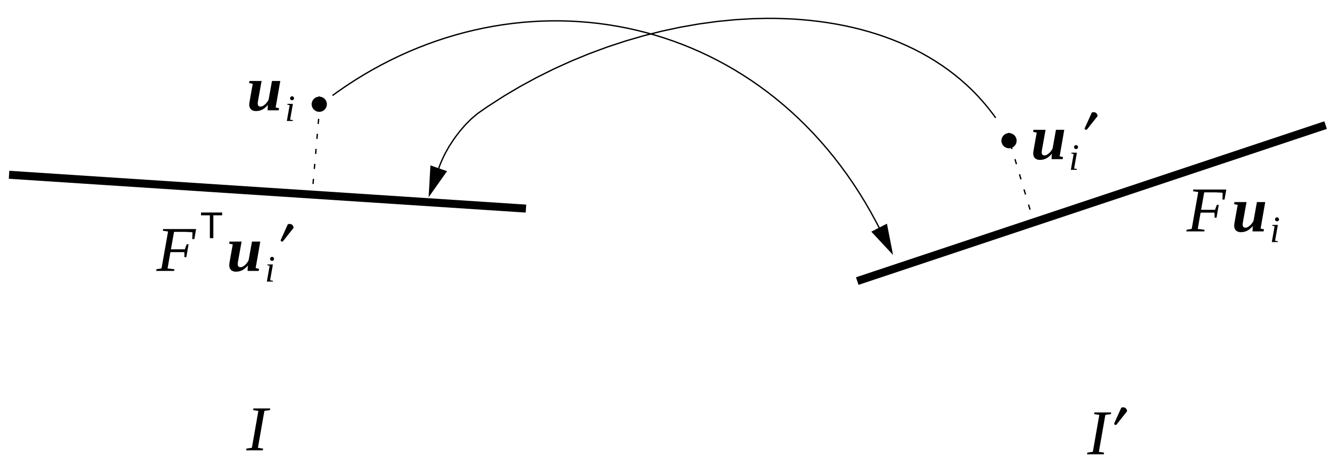
E.g., Levenberg-Marquardt.

Slower than linear algorithms, but may give better results.

A number of geometric error functions (and linear approximations to these) have been defined, including:

- ✦ Symmetric transfer error,
- ✦ Reprojection error.

Symmetric Transfer Error

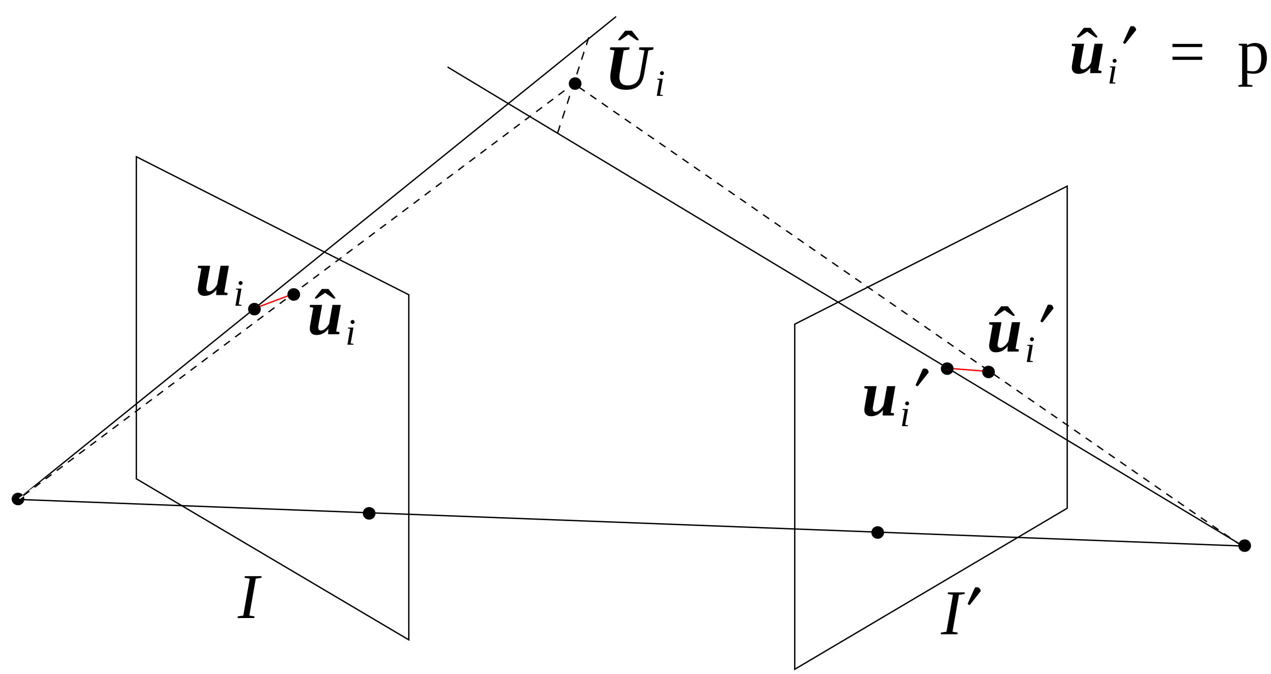


$$\min_F \sum_i \left(d^2(\boldsymbol{u}_i', F\boldsymbol{u}_i) + d^2(\boldsymbol{u}_i, F^\top \boldsymbol{u}_i') \right)$$

Reprojection Error

$$\hat{\boldsymbol{u}}_i = \mathrm{proj}_I(F, \boldsymbol{u}_i, \boldsymbol{u}_i')$$

$$\hat{\boldsymbol{u}}_i' = \mathrm{proj}_{I'}(F, \boldsymbol{u}_i, \boldsymbol{u}_i')$$



$$\min_{F, \hat{\boldsymbol{u}}_i, \hat{\boldsymbol{u}}_i'} \sum_i \left(d^2(\boldsymbol{u}_i, \hat{\boldsymbol{u}}_i) + d^2(\boldsymbol{u}_i', \hat{\boldsymbol{u}}_i') \right)$$

Least Squares Methods

Minimize a sum of squared residuals:

$$\min_{a_1, \ldots, a_M} \sum_{i=1}^{n} \left(\frac{y_i - y(x_i; a_1, \ldots, a_M)}{\sigma_i} \right)^2 .$$

This is optimal (in terms of maximum likelihood) when the measurements y_i are normally distributed about the true values $y(x_i)$.

A single model is required to be fit to the *entire* data set.

Often, this is not a accurate model of real data.

Line Fitting Example

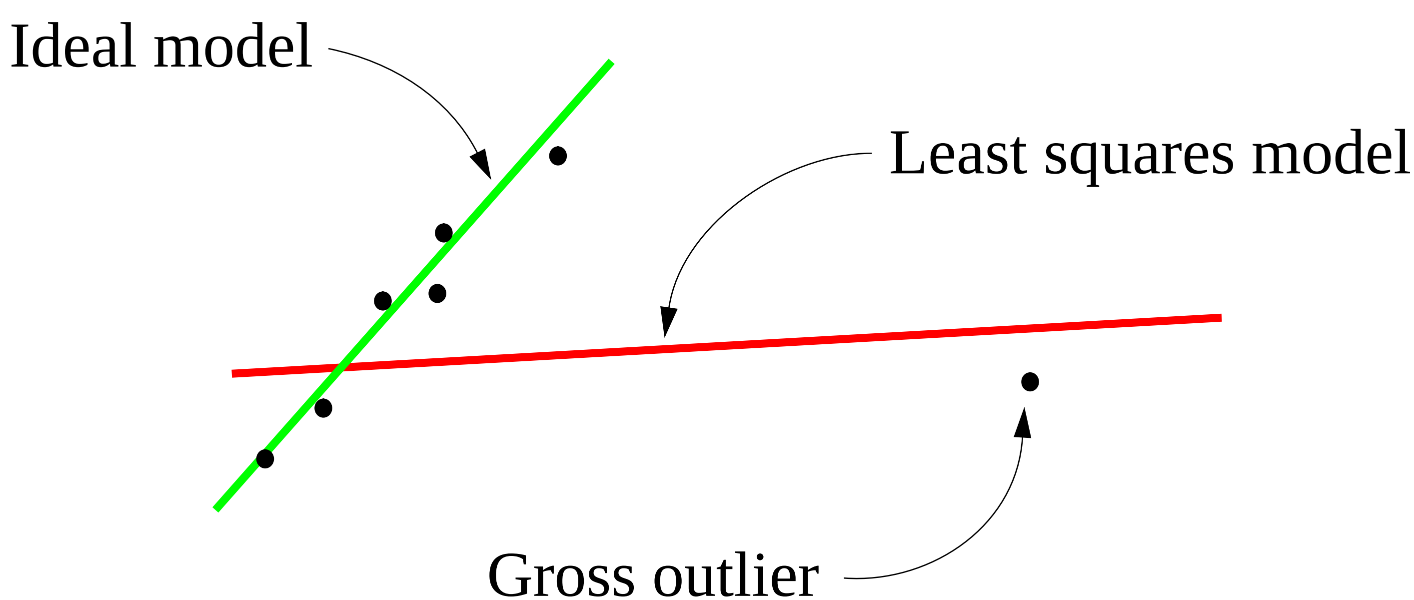


Image Correspondence Data

Generated by hand or by automated matching methods that employ heuristics.

Two types of errors:

Bad locations: The corresponding point is correct, but its location is in error by a small amount (e.g., less than 2σ).

False matches: The point is matched to the wrong point. The error in the location can be large.

We want to use estimators that are robust to both types of errors.

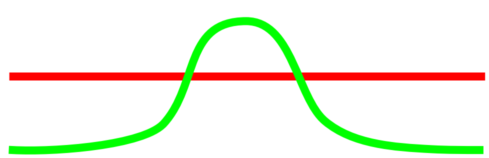
Synthetic Data

Inlying correspondences:

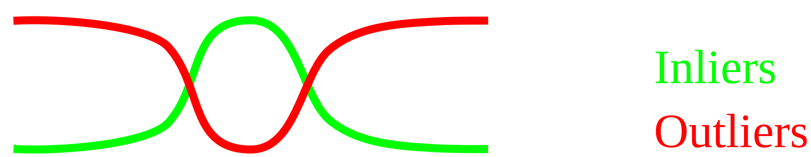
Errors are normally distributed with mean zero, $0 \leq \sigma \leq 3$ pixels.

Outlying correspondences:

Errors are uniformly distributed between -15 and 15 pixels.

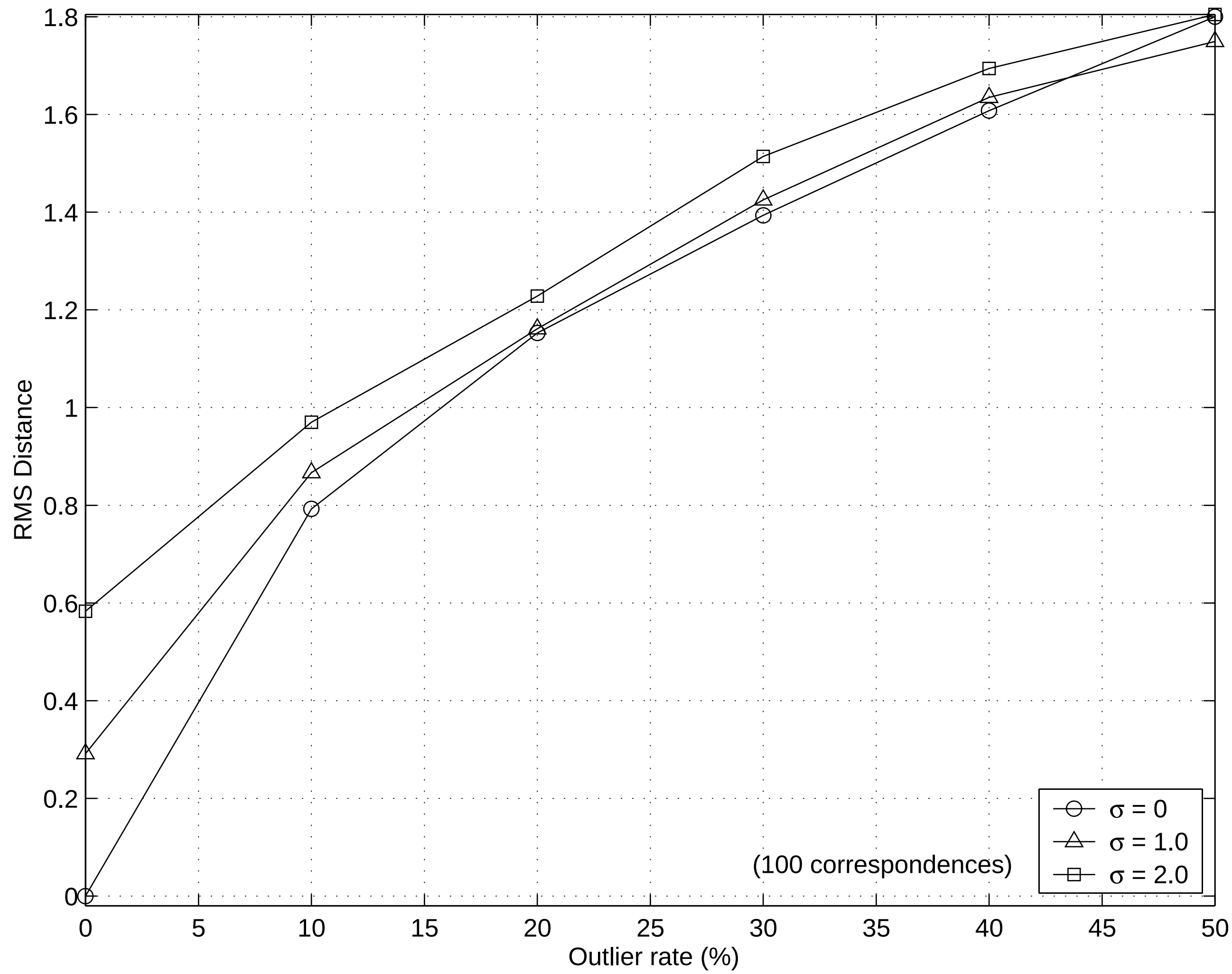


Current distributions



Better distributions

Performance of 8-Pnt. Alg. w/ Outliers



Robust Estimators

Robust: The estimate is insensitive to small deviations of the data from the model for which the estimator has been designed.

Small departures:

- ✦ Small deviations for any number of data points.

  $\Rightarrow$ **Inliers**

- ✦ Large deviations for a small number of data points.

  $\Rightarrow$ **Outliers**

Robust Methods

- M-Estimators
- Least Median of Squares (LMS)
- Random Sample Consensus (RANSAC)

M-Estimators

The M-estimates method tries to reduce the effect of outliers by replacing the squared residuals by a function of the residuals that is symmetric, positive, has a unique minimum at zero, and is less increasing than square.

Let r_i be the residual of the ith correspondence. E.g., $r_i = \boldsymbol{u}_i'^{\top} F \boldsymbol{u}_i$.

Standard least squares: Minimize $\sum_i r_i^2$.

M-Estimator: Minimize $\sum_i w(r_i) r_i^2$.

Example weight function: $$w(r) = \begin{cases} 1 & |r| \le \sigma \\ \dfrac{\sigma}{|r|} & \sigma < |r| < 3\sigma \\ 0 & 3\sigma < |r| \end{cases}$$

Least Median of Squares

Find the fundamental matrix F that minimizes the median of the squared residuals computed for the entire set of correspondences.

Find F that minimizes

$$\operatorname*{med}_i \; r_i^2(F)$$

where

$$r_i^2(F) \;=\; d^2(\boldsymbol{u}_i', F\boldsymbol{u}_i) + d^2(\boldsymbol{u}_i, F^\top \boldsymbol{u}_i')$$

is the residual for the i^{th} correspondence $\boldsymbol{u}_i \leftrightarrow \boldsymbol{u}_i'$.

The matrix F calculated in this way is robust to a moderate number of incorrectly corresponded points.

Least Median of Squares

- Search through the space of possible fundamental matrices.

- Potential fundamental matrices are generated by applying a quick (linear) algorithm to small subsamples of the correspondence set.

- Because this space is very large, only randomly chosen subsets of the correspondences are examined.

Least Median of Squares

for k = 1 to M do

begin

1. Choose s correspondences from $\{(\boldsymbol{u}_i, \boldsymbol{u}_i'), i = 1, \ldots, n\}$.
   Let C_k be this set of correspondences.

2. Compute F_k from C_k.

3. Compute $med_k = \underset{1 \le i \le n}{\text{med}}\ r_i^2(F_k)$.

4. Set med_{min} = min(med_{min}, med_k) and let F_{min} be the corresponding F_k.

end

$\sigma_{robust} = 1.4826[1 + 5/(n-s)]\sqrt{med_{min}}$.

Calculate F using all $(\boldsymbol{u}_i, \boldsymbol{u}_i')$ for which $r_i \le \sigma_{robust}$.

Parameters for LMS Algorithm

Let α be ratio of outliers in the data.

How many subsamples, M, of $\{(\boldsymbol{u}_i, \boldsymbol{u}_i')\}$, taken s at a time, should be examined?

Want P[at least one of the M samples includes only inliers] $\geq p$.

Typically, $p = 0.99$.

Probability at least one sample is good is:

$$p = 1 - [1 - (1 - \alpha)^s]^M .$$

Solve for M: $M = \dfrac{\log(1 - p)}{\log(1 - (1 - \alpha)^s)}$.

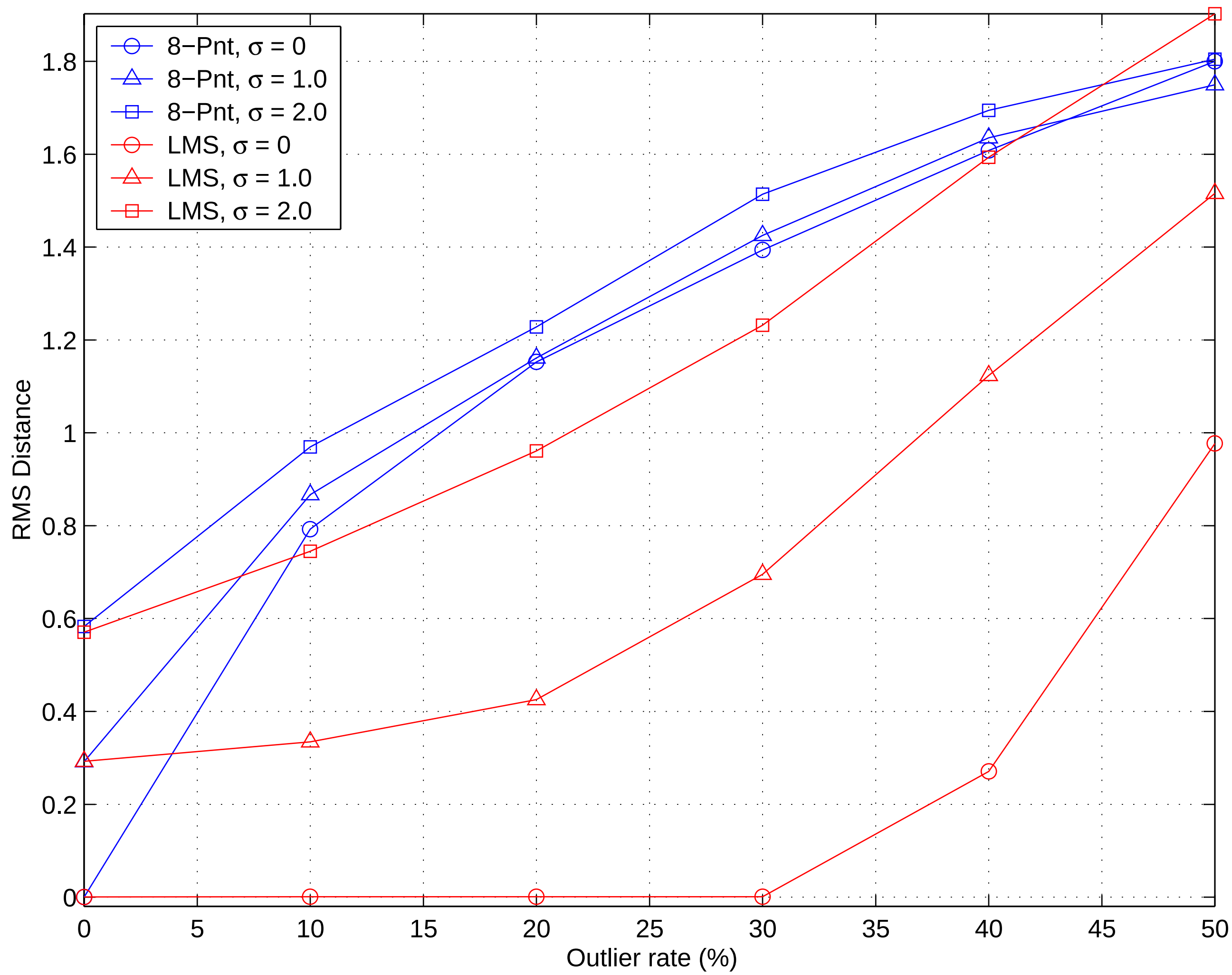
Parameters for LMS Algorithm

How large should each subsample be?

Number of samples, M, to ensure, with a probability $p = 0.99$, that at least one sample has no outliers, for a given s and α.

Sample size	Proportion of outliers, α						
s	5%	10%	20%	25%	30%	40%	50%
2	2	3	5	6	7	11	17
3	3	4	7	9	11	19	35
4	3	5	9	13	17	34	72
5	4	6	12	17	26	57	146
6	4	7	16	24	37	97	293
7	4	8	20	33	54	163	588
8	5	9	26	44	78	272	1177

Performance of LMS Algorithm



Uncertainty in the Fundamental Matrix

Covariance matrix:

$$\Sigma_F = E[(F - \bar{F})(F - \bar{F})^{\mathsf{T}}]$$ where F is interpreted as a 9-vector.

Statistical estimation of Σ_F:

$$E_n[F] = \frac{1}{n} \sum_{i=1}^{n} F_i$$

$$\tilde{\Sigma}_F = \frac{1}{n-1} \sum_{i=1}^{n} [(F_i - E_n[F])(F_i - E_n[F])^{\mathsf{T}}]$$

Analytical estimation of Σ_F

- Uses linear approximations of nonlinear functions (via Taylor series).
- Assumes measurement errors are independent from one point to the next.

$$\Sigma_F \approx \frac{2S}{n-7} H^{-\top}$$

where:

S is the value of the criterion function (e.g., $\sum_i (\boldsymbol{u}_i'^{\top} F \boldsymbol{u}_i)^2$) at the minimum.

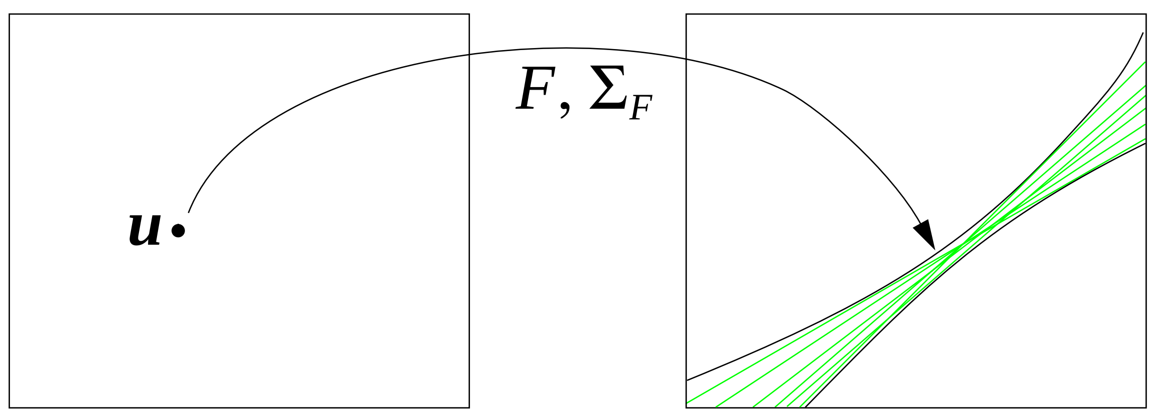
H is the Hessian of the criterion function, which can be obtained as a by-product of a nonlinear minimization.

Using Σ_F

Σ_F can be used to:

- ✔ Bound the locations of epipolar lines,
- ✔ Compute the uncertainty in projective reconstruction,
- ✔ Improve camera calibration.

Epipolar bands



3D Scene Representation as a Collection of Images and Fundamental Matrices

Given:

1. A set of images of some 3D scene,

2. the fundamental matrices relating those images, and

3. a set of corresponding points in the images

Generate new views of the 3D scene *without* reconstructing the 3D scene.

Assumption

- We don’t necessarily known the internal and external parameters of the cameras.

When cameras are **weakly calibrated** (only F is known, not the internal camera parameters), then we can only generate images that are accurate up to an unknown projective transformation.

The 3D Reconstruction Method

1. Calibrate the system of cameras.

2. Determine image correspondences.

3. Fuse (triangulate) corresponding points into 3D points.

4. Approximate the set of 3D points by a surface. *(tricky)*

5. Project this surface into the new image.

Image-based Method

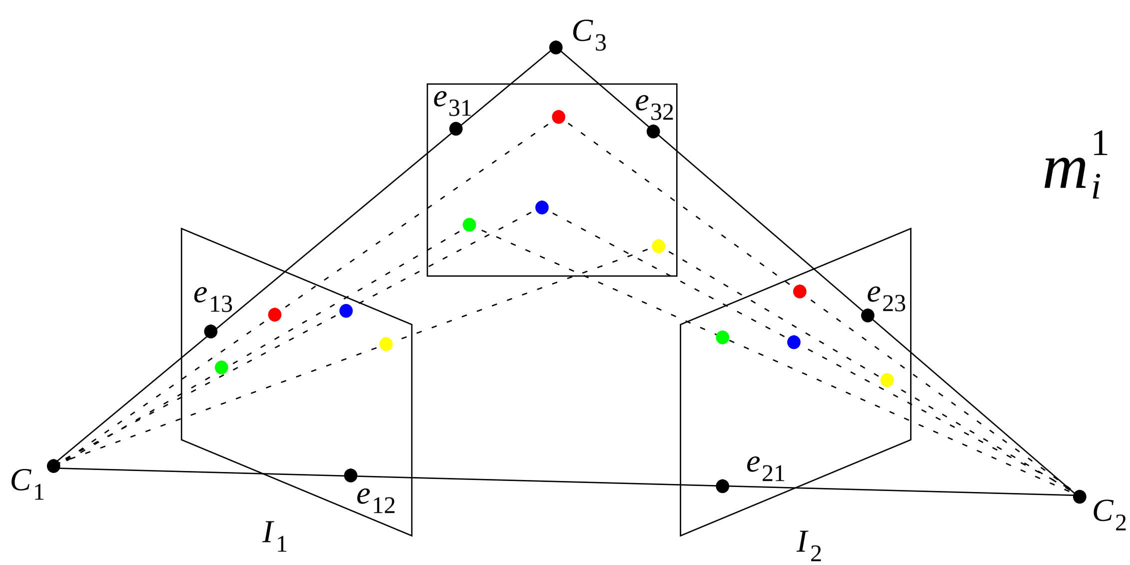
A large number of views of a rigid scene under rigid transformations can be represented by a combination of a small number of views of the scene.

All calculations are done in image planes, none in 3-space.

Defining the Third View

The **center of projection** of the third camera is defined by the epipoles of that camera in the two reference images.

The **image plane** of the third camera is defined up to a projective transformation by the images of four corresponding points in the two reference images. (These must satisfy epipolar constraint of *F*.)



$$m_i^1 = H m_i^2$$

To Construct a New Image

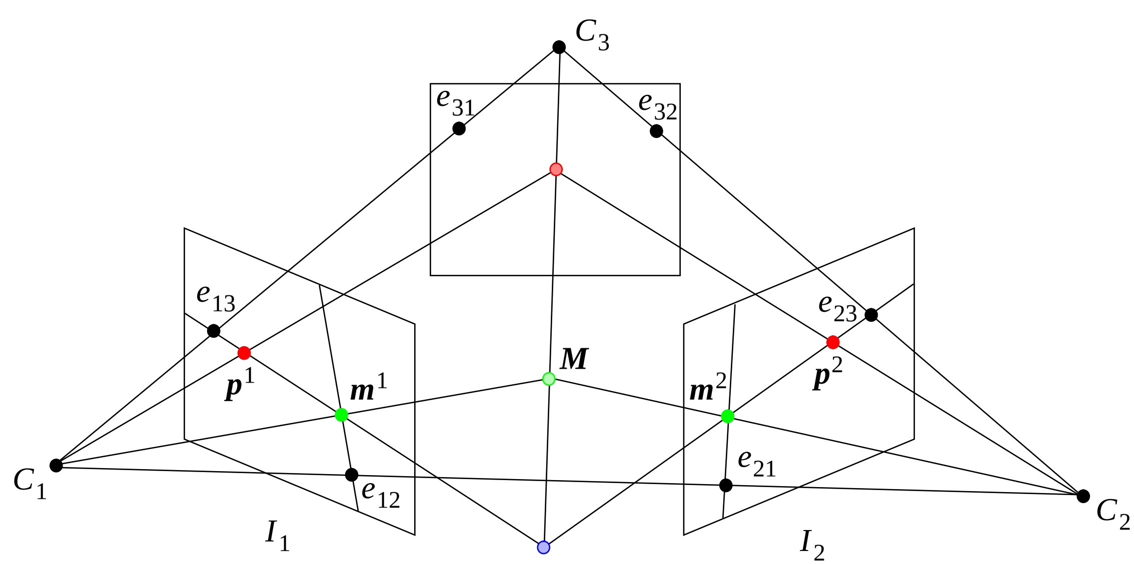
- Need a set of point correspondences between the two reference images.

  The denser the better, since only these points will be imaged.

- Need a way of to project virtual scene points (defined only by their images in the reference images) onto the virtual image plane.

Projection of Virtual Scene Points

$$\boldsymbol{p}^1 = (H^{-\mathsf{T}}(\boldsymbol{e}_{23} \times \boldsymbol{m}^2)) \times (\boldsymbol{e}_{13} \times \boldsymbol{m}^1)$$
$$\boldsymbol{p}^2 = (H^{\mathsf{T}}(\boldsymbol{e}_{13} \times \boldsymbol{m}^1)) \times (\boldsymbol{e}_{23} \times \boldsymbol{m}^2)$$



References

1. Z. Zhang. Parameter Estimation Techniques: A Tutorial with Application to Conic Fitting. *Image and Vision Computing.* Vol. 15 (1997), pp. 59-76.

2. Z. Zhang. Determining the Epipolar Geometry and its Uncertainty: A Review. *Intl. Journal of Computer Vision*, 27(2), pp. 161-195, 1998. Also INRIA Research Report No. 2927.

3. Q.-T. Luong, R. Deriche, O.D. Faugeras, and T. Papadopoulo. On Determining the Fundamental Matrix: Analysis of Different Methods and Experimental Results. INRIA Technical Report No. 1894, April 1993.

4. R.I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

5. R.I. Hartley. In Defence of the 8-Point Algorithm. *IEEE Trans. Pattern Analysis and Machine Vision*. Vol. 19, No. 6, June 1997.

6. O. Faugeras and S. Laveau. Representing Three-Dimensional Data as a Collection of Images and Fundamental Matrices for Image Synthesis. *Proceedings of ICPR94*, pp. 689-691, 1994. Also INRIA Research Report No. 2205.

7. S. Seitz and C. Dyer. Toward Image-Based Scene Representation Using View Morphing, *Proc. 13th Intl. Conf. on Pattern Recognition*, 1996, pp. 84-89.