Epipolar Geometry

Goal: Explore basic geometry of multiple images to support the inference of 3D scene structure from two or more images of the scene.

Consider two perspective images of a scene as taken from a stereo pair of cameras (or equivalently, assume the scene is rigid and imaged with a single camera from two different locations).

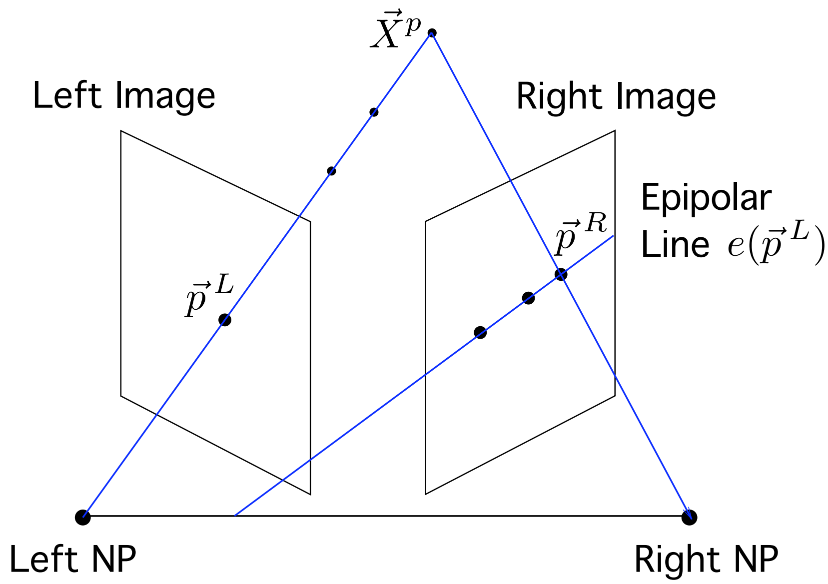
- Given a scene point $\vec{X}^p$ which is imaged in the "left" camera at $\vec{p}^{\,L}$, where could the image of the same point be in the right camera?
- The relationship between such *corresponding image points* turns out to be both simple and useful; i.e., the corresponding point in the "right" camera, $\vec{p}^{\,R}$, is constrained to lie on a line.

Demos: 3dRecon/grappleFmatrix.m (utvis)

Readings: See Sections 10.1 and 15.6 of Forsyth and Ponce.

Epipolar Line

Perspective Cameras: Let $\vec{d}^{\,L}$ and $\vec{d}^{\,R}$ be the 3D nodal points of the left and right cameras. The projection of scene point $\vec{X}^p$ onto the left image, $\vec{p}^{\,L}$, is the intersection of the image plane with the line through $\vec{X}^p$ and $\vec{d}^{\,L}$.



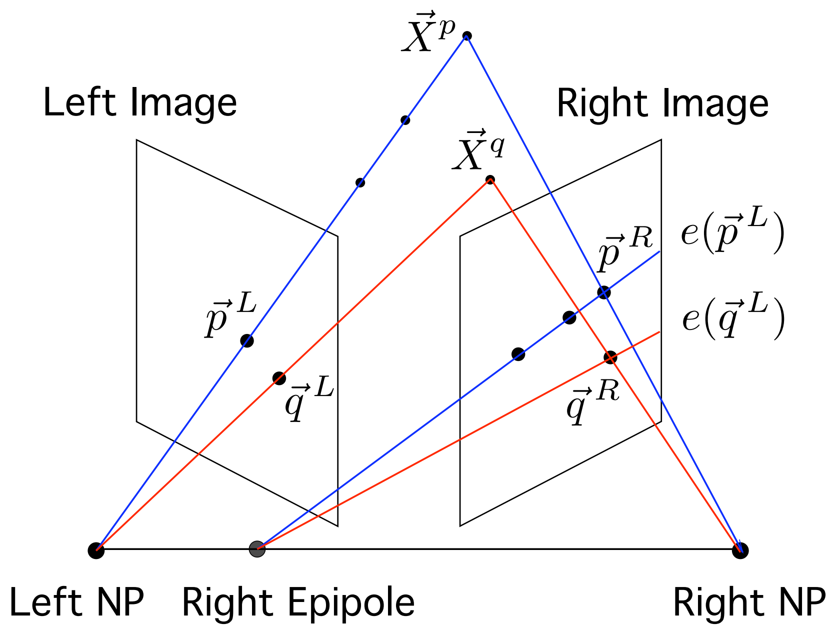
Epipolar Line: The scene point $\vec{X}^p$ that projects to $\vec{p}^{\,L}$ lies on a ray from $\vec{d}^{\,L}$ through $\vec{p}^{\,L}$, but the position of $\vec{X}^p$ on the ray is unknown. Importantly, the image of the ray can be shown to be a line in the right image, called the *epipolar line*, $e(\vec{p}^{\,L})$.

Epipolar Plane: Consider the 3D plane defined by $\vec{p}^{\,L}$ and the nodal points, $\vec{d}^{\,L}$ and $\vec{d}^{\,R}$:

- This *epipolar plane* also contains the 3D scene point $\vec{X}^p$.
- It must also contain the image of $\vec{X}^p$ in the right camera, i.e., $\vec{p}^{\,R}$.
- The image of the ray from $\vec{d}^{\,L}$ through $\vec{p}^{\,L}$ is the line formed by the intersection of the right image plane and the epipolar plane.

Epipolar Constraint

Epipolar Constraint: Let $\vec{p}^{\,L}$ be the left image point for the 3D scene point $\vec{X}^p$. Then, the corresponding point in the right image, $\vec{p}^{\,R}$, lies on the epipolar line $e(\vec{p}^{\,L})$.



The epipolar line $e(\vec{p}^{\,L})$ depends on the position of the left image point. A different point, $\vec{q}^{\,L}$, generally produces a different epipolar line, $e(\vec{q}^{\,L})$.

Epipole: All epipolar lines in the right image pass through a single point (possibly at infinity), called the *right epipole*. It is the intersection of the line containing the nodal points, $\vec{d}^{\,L}$ and $\vec{d}^{\,R}$, with the right image plane. The line through the two nodal points must be in all the epipolar planes, and hence its image must be on all the epipolar lines.

Constraints on Correspondence

Clearly we can swap the labels “left” and “right” in the above analysis, it does not matter which image we start with.

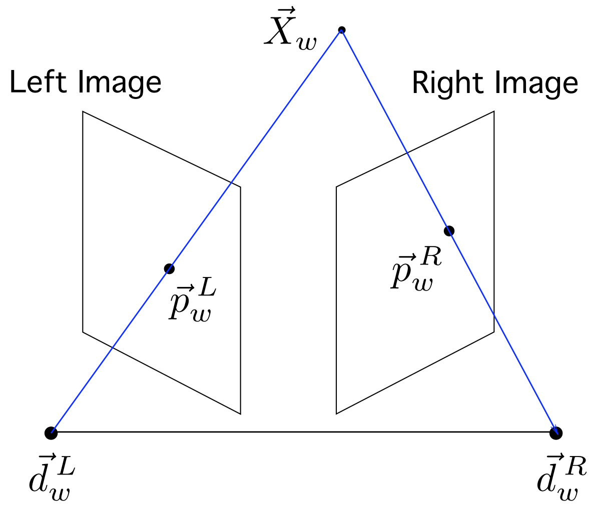
The previous analysis showed there is a mapping between points in one image and epipolar lines in the other. This mapping is computationally useful as it provides strong constraints on corresponding points in two images of the same scene.

- E.g., for each point in one image, we could limit the search for a corresponding point in the second image to just the epipolar line (instead of searching the whole second image). Searching for such corresponding points is central to stereo depth estimation.
- Alternatively, given a set of hypothesized correspondences, we can use the epipolar constraints to identify (some) outliers.

For such applications we need to be able to estimate the parameters the mapping from points to epipolar lines. We next consider this, first with two calibrated cameras, and then in the uncalibrated case.

Epipolar Plane

Consider a 3D point $\vec{X}_w$, in world coordinates, its image in the left and right image planes, $\vec{p}_w^{\,L}$ and $\vec{p}_w^{\,R}$, and nodal points, $\vec{d}_w^{\,L}$ and $\vec{d}_w^{\,R}$.



The 3D plane defined by $\vec{p}_w^{\,R}$, $\vec{d}_w^{\,L}$ and $\vec{d}_w^{\,R}$, has normal direction

$$(\vec{d}_w^{\,L} - \vec{d}_w^{\,R}) \times (\vec{p}_w^{\,R} - \vec{d}_w^{\,R}) \;, \tag{1}$$

where '$\times$' denotes the vector cross-product.

The left image point $\vec{p}_w^{\,L}$ must also lie in the plane. So the vector from $\vec{p}_w^{\,L}$ to $\vec{d}_w^{\,L}$ must be perpendicular to the normal; i.e.,

$$(\vec{p}_w^{\,L} - \vec{d}_w^{\,L})^T \left[(\vec{d}_w^{\,L} - \vec{d}_w^{\,R}) \times (\vec{p}_w^{\,R} - \vec{d}_w^{\,R}) \right] \;=\; 0 \;. \tag{2}$$

The epipolar plane equation is important as it provides constraints on the left and right image locations to which a 3D scene point projects. We formulate this next.

Image Formation

Let left and right camera-centered coordinate frames have origins at the respective nodal points, and z-axes aligned with the optical axes.

A 3D point, $\vec{X}_w$ is transformed into the left coordinate frame using the external parameters of the left camera (see the image formation notes):

$$\vec{X}_c^L \;=\; M_{ex}^L \begin{pmatrix} \vec{X}_w \\ 1 \end{pmatrix} \,, \tag{3}$$

with M_{ex}^L the 3×4 matrix $M_{ex}^L \;=\; [R^L \;\; -R^L \vec{d}_w^L]$; here, R^L is a 3×3 rotation matrix, and $\vec{d}_w^L$ is the left nodal point in world coordinates.

The left image of $\vec{X}_c^L$ is given by perspective projection,

$$\vec{p}_c^L \;=\; \frac{f^L}{X_{3,c}^L}\vec{X}_c^L \;=\; \begin{pmatrix} p_{1,c}^L \\ p_{2,c}^L \\ f^L \end{pmatrix} \,, \tag{4}$$

where, f^L is the nodal distance.

Finally, given (3) above, the camera and world coordinates of 3D points on the left image plane are related by

$$\vec{p}_c^L \;=\; M_{ex}^L \begin{pmatrix} \vec{p}_w^L \\ 1 \end{pmatrix} \;=\; R^L(\vec{p}_w^L - \vec{d}_w^L) \,. \tag{5}$$

And similarly, for image points in the right camera,

$$\vec{p}_c^R \;=\; R^R(\vec{p}_w^R - \vec{d}_w^R) \,. \tag{6}$$

The Essential Matrix

The epipolar plane, on which the left and right projections of $\vec{X}_w$ must lie, is given by

$$(\vec{p}_w^{\,L} - \vec{d}_w^{\,L})^T \left[(\vec{d}_w^{\,L} - \vec{d}_w^{\,R}) \times (\vec{p}_w^{\,R} - \vec{d}_w^{\,R}) \right] \; = \; 0 \, . \tag{7}$$

- Based on (5) and (6), we know that $\vec{p}_w^{\,L} - \vec{d}_w^{\,L} \; = \; (R^L)^T \vec{p}_c^{\,L}$, and $\vec{p}_w^{\,R} - \vec{d}_w^{\,R} = (R^R)^T \vec{p}_c^{\,R}$.
- We can replace the cross-product by the equivalent matrix-vector product,

$$\vec{T} \times \vec{p} \; = \; [\vec{T}]_\times \vec{p}, \quad \text{where } [\vec{T}]_\times \; = \; \begin{pmatrix} 0 & -T_3 & T_2 \\ T_3 & 0 & -T_1 \\ -T_2 & T_1 & 0 \end{pmatrix} .$$

  $[\vec{T}]_\times$ is rank 2, with two identical, non-zero singular values.

It therefore follows that the epipolar constraint in (7) can be re-expressed as

$$(\vec{p}_c^{\,L})^T E \, \vec{p}_c^{\,R} \; = \; 0 \, , \tag{8}$$

where E is the 3×3 *essential matrix* (or E-matrix):

$$E \; = \; R^L \, [\vec{d}_w^{\,L} - \vec{d}_w^{\,R}]_\times \, (R^R)^T \, . \tag{9}$$

Properties of the Essential Matrix

1. Clearly, any nonzero scalar multiple of the E-matrix provides an equivalent epipolar constraint (8).

2. From (9) it follows that the E-matrix has rank 2, with two equal non-zero singular values and one singular value at 0.

3. Given a point $\vec{p}_c^{\,L}$ in the left image, the epipolar constraint (8) states that the corresponding point $\vec{p}_c^{\,R}$ in the right image must be on a line. I.e., with $\vec{a} \;=\; E^T \vec{p}_c^{\,L}$, it follows that

$$\vec{a}^T \vec{p}_c^{\,R} \;=\; a_1 p_{1,c}^R + a_2 p_{2,c}^R + a_3 f^R \;=\; 0\;.$$

The normal to the line is (a_1, a_2).

4. The right epipole $\vec{e}_c^{\,R}$ lies on all epipolar lines. That is, for all left image points, $\vec{p}_c^{\,L}$, the right epipole must satisfy

$$(\vec{p}_c^{\,L})^T E\, \vec{e}_c^{\,R} \;=\; 0\;. \tag{10}$$

It follows that $\vec{e}_c^{\,R}$ must be a null vector for E. Further, using

$$(\vec{p}_c^{\,L})^T E \vec{e}_c^{\,R} \;=\; \alpha (\vec{p}_c^{\,L})^T R^L [\vec{d}_w^{\,L} - \vec{d}_w^{\,R}]_\times (R^R)^T R^R (\vec{d}_w^{\,L} - \vec{d}_w^{\,R})$$

we find that the epipole can be written as

$$\vec{e}_c^{\,R} \;=\; \alpha M_{ex}^R \begin{pmatrix} \vec{d}_w^{\,L} \\ 1 \end{pmatrix} \;=\; \alpha\, R^R\, (\vec{d}_w^{\,L} - \vec{d}_w^{\,R})\;.$$

For a point $\vec{p}_c^{\,R}$ in the right image, analogous expressions provide the epipolar line in the left image, and the left epipole.

Internal Calibration

We wish to rewrite the epipolar constraint (8) in terms of homogeneous pixel coordinates $\vec{x}^{\,L} = (x^L, y^L, 1)^T$, where (x^L, y^L) are the coordinates of an image point in terms of pixels.

The internal calibration matrix M_{in}^L provides the transformation from camera coordinates to homogeneous pixel coordinates (see the image formation notes),

$$\vec{x}^{\,L} \;=\; M_{in}^L \vec{p}_c^{\,L} \; . \tag{11}$$

For example, a camera with rectangular pixels of size $1/s_x$ by $1/s_y$, with nodal distance f, and piercing point (o_x, o_y) (i.e., the intersection of the optical axis with the image plane provided in pixel coordinates) has the internal calibration matrix

$$M_{in} \;=\; \begin{pmatrix} s_x & 0 & o_x/f \\ 0 & s_y & o_y/f \\ 0 & 0 & 1/f \end{pmatrix} . \tag{12}$$

We can use (11) to rewrite the epipolar constraint in terms of pixel coordinates.

The Fundamental Matrix

Using (11) we can rewrite the epipolar constraint (8) for *homogeneous* pixel coordinates in the left and right images; i.e.,

$$(\vec{x}^{\,L})^T F \,\vec{x}^{\,R} \;=\; 0\,. \tag{13}$$

Here the *fundamental matrix* (or F-matrix) is given by

$$F \;=\; (M_{in}^L)^{-T}\, E\,(M_{in}^R)^{-1}\,, \tag{14}$$

where the notation M^{-T} denotes the transpose of the inverse of M.

Like the E-matrix, the F-matrix has rank 2, but the two nonzero singular values need not be equal. The over-all scale of the F-matrix does not effect the epipolar constraint (13). So there remain 7 degrees of freedom in F.

The right (left) null vector of F gives the homogeneous pixel coordinates for the right (left, resp.) epipole.

More explicitly, for example, the epipolar constraint (13) states that, given a point $\vec{x}^{\,L}$ in the left image, the corresponding point $\vec{x}^{\,R}$ in the right image must be on the epipolar line

$$\vec{a}^{\,T}\vec{x}^{\,R} \;=\; a_1\, x^R + a_2\, y^R + a_3 \;=\; 0,$$

where $\vec{a} = F^T\vec{x}^{\,L}$.

Estimating the Fundamental Matrix

Given corresponding image points $\{(\vec{x}_k^L, \vec{x}_k^R)\}_{k=1}^K$ we wish to estimate the F-matrix.

Gold Standard Approach: Suppose the noise in the point positions $\vec{x}_k^\mu$, for $\mu = L, R$ is additive, independent and normally distributed with mean zero and covariance Σ_k^μ:

$$\vec{x}_k^\mu = \vec{m}_k^\mu + \vec{n}_k^\mu, \tag{15}$$

where $\vec{m}_k^\mu$ is the true position and $\vec{n}_k^\mu$ is the noise. Note: there is no noise in the third component of $\vec{x}_k^\mu$ in homogeneous coordinates.

Then the (maximum likelihood) problem is to find $F \in \Re^{3\times 3}$ and $\vec{m}_k^\mu$, for $k = 1, \ldots, K$ and $\mu = L, R$, such that the following objective function is minimized:

$$\mathcal{O} \equiv \sum_{\mu\in\{L,R\}} \sum_{k=1}^{K} (\vec{x}_k^\mu - \vec{m}_k^\mu)^T (\Sigma_k^\mu)^\dagger (\vec{x}_k^\mu - \vec{m}_k^\mu) \tag{16}$$

where $(\Sigma_k^\mu)^\dagger$ denotes the pseudo-inverse. We minimize this objective function $\mathcal{O}$ subject to the epipolar constraints:

$$(\vec{m}_k^L)^T F \vec{m}_k^R = 0, \quad k = 1, \ldots, K, \tag{17}$$

$$\mathrm{rank(F)} = 2 \tag{18}$$

Thus, $\mathcal{O}$ is a quadratic objective function for the $\vec{m}_k^\mu$'s, with nonlinear constraints (17) and (18).

Alternative Estimation Approaches

We would like to avoid a nonlinear optimization problem, at the risk of a noiser estimate of the F-matrix than that provided by the gold standard objective. One simplification is to ignore the noise in $\vec{x}_k^L$ when estimating the epipolar line $e(\vec{m}_k^L)$. That is, let's say the corresponding right point $\vec{x}_k^R$ should be close to $e(\vec{x}_k^L)$ instead of $e(\vec{m}_k^L)$.

This epipolar line $e(\vec{x}_k^L)$ can be written as

$$(\vec{n}^T, c)\,\vec{x}^R = 0\,, \tag{19}$$

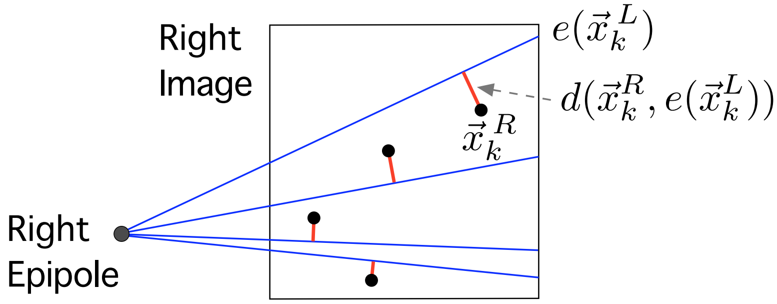
where

$$\begin{pmatrix} \vec{n} \\ c \end{pmatrix} = \frac{1}{||(I_2\ \vec{0})F^T\vec{x}_k^L||_2}\,F^T\,\vec{x}_k^L. \tag{20}$$

The normalization in (20) ensures $\vec{n}$ is the *unit normal* to $e(\vec{x}_k^L)$. Then,

$$d(\vec{x}_k^R, e(\vec{x}_k^L)) \equiv (\vec{n}^T, c)\,\vec{x}^R\,, \tag{21}$$

is the perpendicular distance between $\vec{x}_k^R$ and the epipolar line $e(\vec{x}_k^L)$.



We could minimize the sum of squared epipolar distances $d(\vec{x}_k^R, e(\vec{x}_k^L))$ for $k = 1 \ldots K$. However, due to the normalization factor in (20), the objective function is not quadratic in the unknown F.

Algebraic Error

Instead, consider the reweighted epipolar distance objective function

$$
\begin{aligned}
\mathcal{O}(F) &\equiv \sum_{k=1}^{K} w(\vec{x}_k^{\,L})\, d^2(\vec{x}_k^{\,R}, e(\vec{x}_k^{\,L})) \\
&= \sum_{k=1}^{K} \left[\, (\vec{x}_k^{\,L})^T\, F\, \vec{x}_k^{\,R} \,\right]^2 . \qquad (22)
\end{aligned}
$$

Here the weights $w(\vec{x}_k^{\,L})$ are chosen to provide a quadratic objective function $\mathcal{O}(F)$. That is,

$$
w(\vec{x}_k^{\,L}) \;=\; ||\, (I_2\ \vec{0})\, F^T\, \vec{x}_k^{\,L}\, ||_2^2 \, . \qquad (23)
$$

This objective function corresponds to the *algebraic error* in the noise-less epipolar constraint (13).

Equation (22) is a suitable ML estimator when errors in the algebraic constraints (13) are mean zero with constant variance. If the variances differ significantly, we will get poor estimates for F.

Indeed, without any rescaling (which we discuss next), this approach provides excessively noisy estimates of F.

Renormalized 8-Point Algorithm

Hartley (PAMI, 1997) introduced the following algorithm. Given corresponding points $\{(\vec{x}_k^L, \vec{x}_k^R)\}_{k=1}^K$ with $K \geq 8$,

1. Recenter and rescale the image points using M^μ, $\mu = L, R$, such that
$$M^\mu = \begin{pmatrix} s^\mu & 0 & b_1^\mu \\ 0 & s^\mu & b_2^\mu \\ 0 & 0 & 1 \end{pmatrix}, \tag{24}$$
with
$$\frac{1}{K}\sum_{k=1}^{K} M^\mu \vec{x}_k^\mu = (0,0,1)^T, \tag{25}$$
$$\frac{1}{K}\sum_{k=1}^{K} [\, M^\mu \vec{x}_k^\mu - (0,0,1)^T\,]_*^2 = (\sigma_1^2, \sigma_2^2, 0)^T, \tag{26}$$
where $\sigma_1^2 + \sigma_2^2 = 2$. Here $[...]_*^2$ denotes the square of each element. Define $\vec{r}_k^\mu = M^\mu \vec{x}_k^\mu$ for $k = 1, \ldots, K$ and $\mu = L, R$.

2. Minimize the objective function $\mathcal{O}(\hat{F})$
$$\mathcal{O}(\hat{F}) \equiv \sum_{k=1}^{K} \left[(\vec{r}_k^L)^T \, \hat{F} \, \vec{r}_k^R \right]^2. \tag{27}$$
Note this is a linear least squares problem for the elements of $\hat{F}$.
(Continued on next page.)

Renormalized 8-Point Algorithm (Cont.)

3. Project $\hat{F}$ to the nearest rank 2 matrix (with the error measured in the Frobenius norm):
   (a) Form the SVD of $\hat{F} = U\Sigma V^T$. In general $\Sigma = \mathrm{diag}[\sigma_1^2, \sigma_2^2, \sigma_3^2]$ with $\sigma_i^2 \geq \sigma_{i+1}^2$ for $i = 1, 2$.
   (b) Reset $\sigma_3 = 0$.
   (c) Assign $\hat{F}$ to be $U\Sigma V^T$.
4. Undo the normalization of the image points,

$$F = (M^L)^T \hat{F} M^R \tag{28}$$

This algorithm has been found to provide reasonable estimates for the F-matrix given correspondence data with small amounts of noise (see Hartley and Zisserman, 2000).

It is not robust to outliers.

In order to deal with outliers, we apply the Random Sample Consensus (RANSAC) algorithm to the estimation of the F-matrix.

RANSAC Algorithm for the F-Matrix

Suppose we are given corresponding points $\{(\vec{x}_k^L, \vec{x}_k^R)\}_{k=1}^K$, which may include outliers. Let $\epsilon > 0$ be an error tolerance, and T be the number of trials.

Loop T times:

1. Randomly select 8 pairs $(\vec{x}_k^L, \vec{x}_k^R)$.
2. Use the renormalized algorithm to solve for F using only the eight selected pairs of points.
3. Compute perpendicular errors $d(\vec{x}_k^R, e(\vec{x}_k^L))$ and $d(\vec{x}_k^L, e(\vec{x}_k^R))$, see (20) and (21) for $1 \le k \le K$.
4. Identify inliers:

$$\mathrm{In} = \{k : d(\vec{x}_k^L, e(\vec{x}_k^R)) < \epsilon \text{ and } d(\vec{x}_k^R, e(\vec{x}_k^L)) < \epsilon,\ 1 \le k \le K\}.$$

5. If the number of inliers, $|\mathrm{In}|$, is the largest seen so far, remember the current estimate of F and the inlier set, In.

End loop

6. Solve for F using all pairs with $k \in \mathrm{In}$ (i.e., all inliers). Re-solve for the inlier set In as done in steps 3 and 4 above.

One can iterate step 6 until the set of inliers In does not change.

RANSAC: How Many Trials?

Suppose our data set consists of a fraction p inliers, and $1 - p$ outliers.

How many trials T should be done so that we can be reasonably confident that at least one sampled data set of size $d = 8$ was all inliers?

The probability of choosing $d = 8$ inliers from such a population is roughly p^d when $K \gg d$ (it is exactly p^d if we sample with replacement). So the probability that a given trial of RANSAC fails to select d inliers is $1 - p^d$. Therefore, the probability that RANSAC failed to have any trial with d inliers is $(1 - p^d)^T$. In other words, the probability P_0 that at least one of the RANSAC trials will be a success is

$$P_0 = 1 - (1 - p^d)^T$$

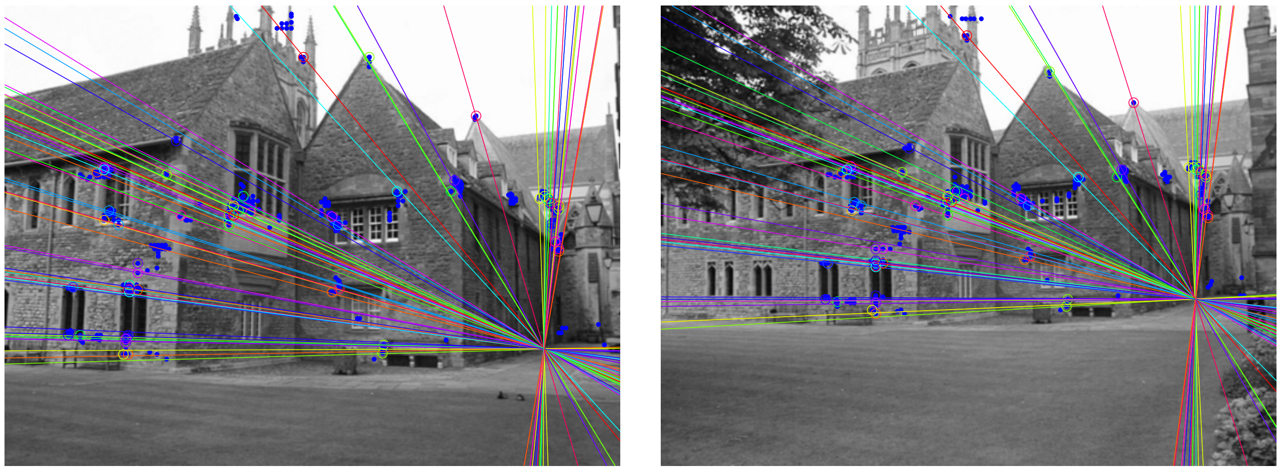
Given an estimate for the fraction of inliers p, and that the probability of at least one successful trial should be P_0 or greater, then we can choose T to be

$$T > \frac{\log(1 - P_0)}{\log(1 - p^d)} .$$

For example, for 70% inliers and $d = 8$, we require $T > 50$. Alternatively, if we only have 50% inliers, the same formula states that T should be chosen to be at least 766.

Example

Given local image features, RANSAC was used to fit the F-matrix.



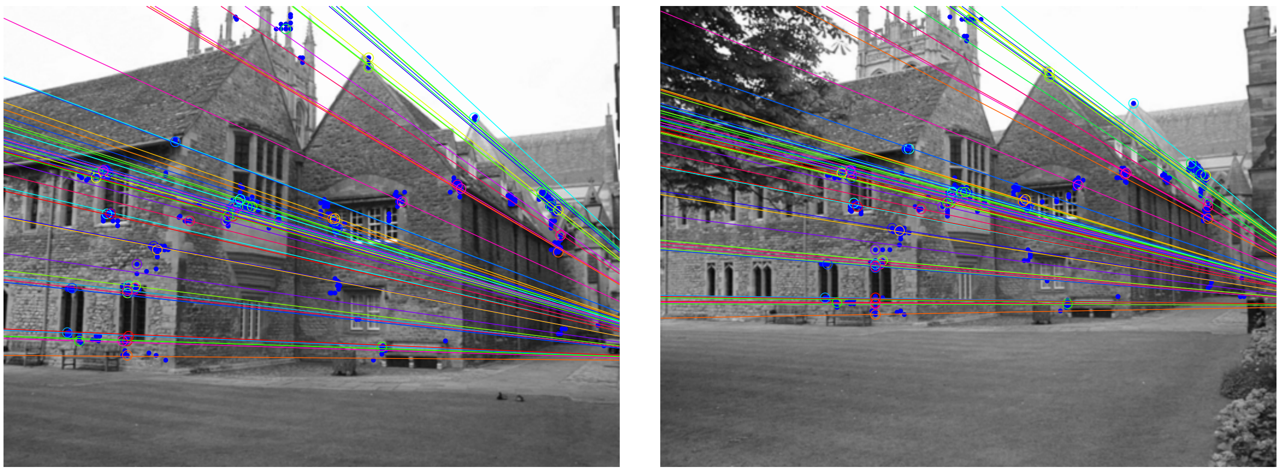
Here have choosen random colours to circle image features. The same colour is then used for the corresponding point in the other image, and also for the epipolar lines generated from these two points.

Note:

1. By construction, each point lies close to the epipolar line generated by its corresponding point in the other image.
2. A visual sanity check can be obtained by sampling other points on one epipolar line, and checking that they also appear somewhere along the corresponding epipolar line. This must be the case since, when the F-matrix is correct, both epipolar lines correspond to the intersection of the scene with the epipolar plane. (Compare the current fit with the result of a poor fit shown on p.19.)
3. The intersection of the epipolar lines corresponds to the epipole in each image. The nodal point of the second camera is on the line (in world coordinates) containing the nodal point of the first camera and the epipole in the first image.

Poorly Fitted F-Matrix

The same local image features were used as in the previous example, and RANSAC was used to fit the F-matrix (but with only 10 trials). The solution it found is displayed below:



Note:

1. The feature points are still near the corresponding epipolar lines. Here 82% of the data points are within 4 pixels of the corresponding epipolar line. In contrast, the solution on the previous page achieved 94%.
2. However, the visual sanity check fails. This is most apparent for (proposed) epipolar planes which intersect the scene over a large range of depths. For example, consider the (proposed) epipolar planes which cut across the tower at the top of the image and at least one of the buildings in front.

Further Readings

R. Hartley (1997) In defense of the eight-point algorithm. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 19(6): 580–593.

R. Hartley and A. Zisserman (2000) *Multiple View Geometry in Computer Vision*. Cambridge University Press.