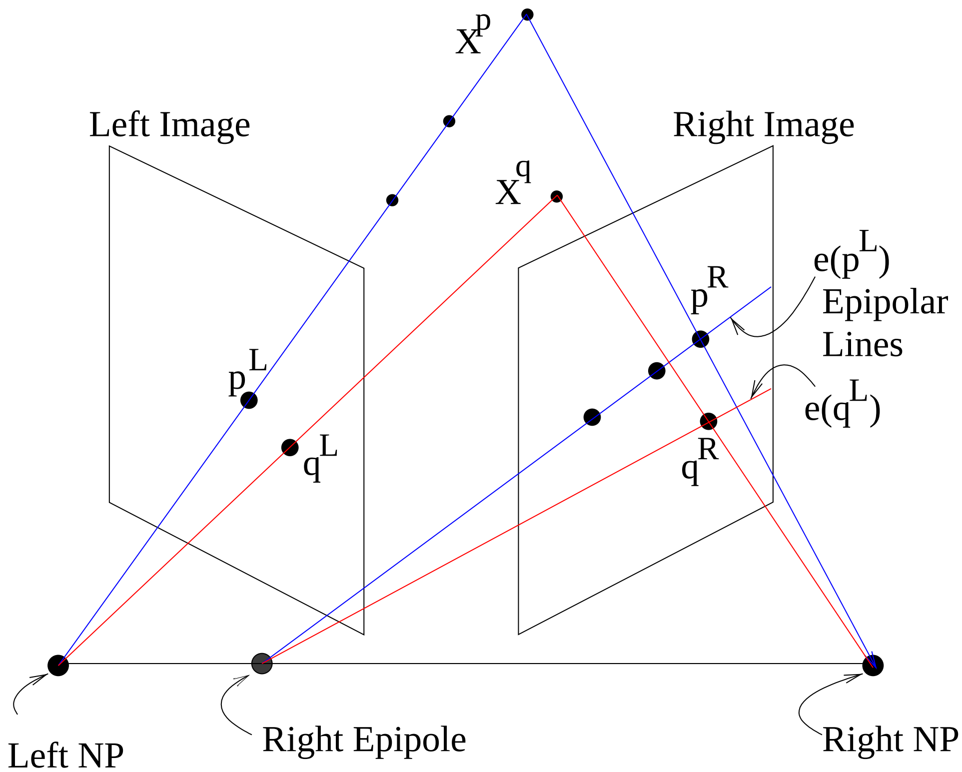
Epipolar Geometry

We 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? We denote this point as $\vec{p}^{\,R}$. The relationship between such *corresponding image points* turns out to be both simple and useful.

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

Epipolar Line

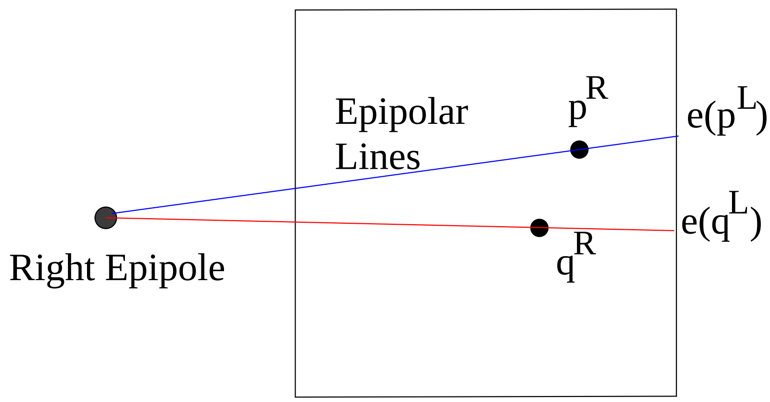
Two Perspective Cameras: Let $\vec{d}^L$ and $\vec{d}^R$ be the 3D positions of the nodal points of the left and right cameras. As discussed earlier in this course, we model perspective projection by placing the image plane in front of the nodal point. A scene point $\vec{X}^p$ is then imaged in the left camera at $\vec{p}^L$, which is the point of intersection of the the image plane with the line containing the scene point $\vec{X}^p$ and the nodal point $\vec{d}^L$ (see previous figure).

Epipolar Line: If we know the left image point $\vec{p}^L$, then the corresponding scene point $\vec{X}^p$ is constrained to be on a ray through this image point. The position of $\vec{X}^p$ on this ray is unknown. However, the image of this whole ray is a line in the right image, namely the epipolar line $e(\vec{p}^L)$.

Epipolar Plane: An alternative geometric view is to consider the 3D plane containing the image point $\vec{p}^L$ along with left and right nodal points $\vec{d}^L$ and $\vec{d}^R$. Then the scene point $\vec{X}^p$ and the corresponding right image point $\vec{p}^R$ must also be on this epipolar plane. The intersection of this epipolar plane with the right image plane provides the epipolar line $e(\vec{p}^L)$.

Epipolar Constraint

Epipolar Constraint: Suppose $\vec{p}^L$ is the left image position for some scene point $\vec{X}^p$. Then the corresponding point $\vec{p}^R$ in the right image must lie on the epipolar line $e(\vec{p}^L)$.



Right Image Plane

Notice the epipolar line $e(\vec{p}^L)$ depends on the position of the point in the left image. For example, another image point $\vec{q}^L$ generally gives rise to a different epipolar line $e(\vec{q}^L)$.

Epipole: All the epipolar lines in the right image pass through a single point (possibly at infinity) called the right epipole. This point is given by the intersection of the line containing the two nodal points $\vec{d}^L$ and $\vec{d}^R$ with the right image plane (see figures above and on p.1). (Notice that the line containing 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. Such a mapping would be computationally useful since it provides strong constraints on corresponding points in two images of the same scene.

- For example, 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 naively searching the whole second image).
- Alternatively, given a set of hypothesized correspondences, we can use the epipolar constraints to identify (some) outliers.

To achieve these applications we need to be able to estimate the mapping from points to epipolar lines, which is what we consider next. We begin with the case of two calibrated cameras, and then consider the uncalibrated case.

Camera Coordinates and Image Formation

Here we apply the image formation model introduced earlier to the stereo imaging setup. We introduce three coordinate frames:

- A world coordinate frame $\vec{X}_w$,
- The left and right camera coordinate frames, $\vec{X}_c^L$ and $\vec{X}_c^R$.

The origins of the left and right camera coordinate frames are at their nodal points, and their z-axes are aligned with the two optical axes. Therefore, as discussed in the image formation notes, given a 3D point $\vec{X}_c^L$ in the left camera's coordinates, the left image of this point is at

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

Here, f^L is the distance between the image plane and the nodal point for the left camera.

A similar expression holds for the location of the right image point. However, before applying this expression, the 3D point $\vec{X}_c^L$ must be transformed from the left to the right camera's coordinates. We do this via the world coordinate frame $\vec{X}_w$.

External Calibration

The external calibration parameters for the left camera provide the 3D rigid coordinate transformation from world coordinates to the left camera's coordinates (see the image formation notes)

$$\vec{X}_c^L = M_{ex}^L[\vec{X}_w^T, 1]^T, \tag{2}$$

with M_{ex}^L a 3×4 matrix of the form

$$M_{ex}^L = \begin{pmatrix} R^L & -R^L\vec{d}_w^L \end{pmatrix}. \tag{3}$$

Here R^L is a 3×3 rotation matrix and $\vec{d}_w^L$ is the location of the nodal point for the left camera in world coordinates.

Similarly, the 3×4 matrix M_{ex}^R provides the external calibration of the right camera.

The Essential Matrix

Let $\vec{p}_w^L$ and $\vec{p}_w^R$ be the left and right image points (written in world coordinates) for some given 3D point $\vec{X}_w$. Then the epipolar constraint states that these two image points and the two nodal points $\vec{d}_w^L$, $\vec{d}_w^R$ are all coplanar. This constraint can be written as

$$(\vec{p}_w^L - \vec{d}_w^L)^T \left[(\vec{d}_w^L - \vec{d}_w^R) \times (\vec{p}_w^R - \vec{d}_w^R)\right] \;=\; 0, \tag{4}$$

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

We rewrite this by replacing the cross-product by an equivalent matrix-vector product,

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

Also, we use (2) and (3) to write $\vec{p}_w^L - \vec{d}_w^L$ in terms of the left camera's coordinates as $\vec{p}_w^L - \vec{d}_w^L \;=\; (R^L)^T \vec{p}_c^L$. Using the analogous expression for the right image point, we find that (4) can be rewritten as

$$(\vec{p}_c^L)^T E \vec{p}_c^R \;=\; 0, \;\; \text{(epipolar constraint)} \tag{5}$$

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

$$E \;=\; R^L [\vec{d}_w^L - \vec{d}_w^R]_\times (R^R)^T. \tag{6}$$

Properties of the Essential Matrix

Clearly, any nonzero scalar multiple of the E-matrix provides an equivalent epipolar constraint (5).

From (6) it follows that the E-matrix has rank 2, with two equal nonzero singular values and one singular value at 0.

Given a point $\vec{p}_c^{\,L}$ in the left image, the epipolar constraint (5) states that the corresponding point $\vec{p}_c^{\,R}$ in the right image must be on the line

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

where $\vec{a} \;=\; E^T \vec{p}_c^{\,L}$.

The right epipole $\vec{e}_c^{\,R}$ is a null vector for E. It can be written as

$$\vec{e}_c^{\,R} \;=\; \alpha M_{ex}^R [(\vec{d}_w^{\,L})^T, 1]^T \;=\; \alpha R^R (\vec{d}_w^{\,L} - \vec{d}_w^{\,R}),$$

where α is a nonzero constant. Notice, using (6),

$$\vec{a}^T \vec{e}_c^{\,R} \;=\; (\vec{p}_c^{\,L})^T E \vec{e}_c^{\,R} \;=\; \alpha (\vec{p}_c^{\,L})^T R^L [\vec{d}_w^{\,L} - \vec{d}_w^{\,R}]_\times (R^R)^T R^R (\vec{d}_w^{\,L} - \vec{d}_w^{\,R}) \;=\; 0,$$

so the epipole is on every epipolar line.

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

Internal Calibration

We wish to rewrite the epipolar constraint (5) 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),

$$\vec{x}^{\,L} = M_{in}^L \vec{p}_c^{\,L}. \tag{7}$$

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

$$M_{in} = \begin{pmatrix} s_x & 0 & o_x/f \\ 0 & s_y & o_y/f \\ 0 & 0 & 1/f \end{pmatrix}. \tag{8}$$

We can use (7) to rewrite the epipolar constraint in terms of pixel coordinates.

The Fundamental Matrix

By using (7) we can rewrite the epipolar constraint (5) in terms of homogeneous pixel coordinates in the left and right images as

$$(\vec{x}^{\,L})^T F \vec{x}^{\,R} \;=\; 0. \tag{9}$$

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

$$F \;=\; (M_{in}^L)^{-T} E (M_{in}^R)^{-1}, \tag{10}$$

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

Similar to 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 (9). So there are 7 remaining 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 (9) 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 independent and normally distributed with mean zero and covariance Σ_k^μ. (Note that there is no noise in the third component of $\vec{x}_k^\mu$.) That is,

$$\vec{x}_k^\mu = \vec{m}_k^\mu + \vec{n}_k^\mu, \tag{11}$$

where $\vec{m}_k^\mu$ is the true position of the point and $\vec{n}_k^\mu$ is the mean zero noise. Then the (maximum likelihood) problem is to find $F \in \Re^{3\times 3}$ along with $\vec{m}_k^\mu$ for $k = 1, \ldots, K$ and $\mu = L, R$, such that the following objective function is minimized:

$$\mathcal{O} \equiv \sum_{\mu \in \{L,R\}} \sum_{k=1}^{K} (\vec{x}_k^\mu - \vec{m}_k^\mu)^T (\Sigma_k^\mu)^\dagger (\vec{x}_k^\mu - \vec{m}_k^\mu) \tag{12}$$

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

$$(\vec{m}_k^L)^T F \vec{m}_k^R = 0, \quad k = 1, \ldots, K, \tag{13}$$

$$\mathrm{rank(F)} = 2 \tag{14}$$

Thus $\mathcal{O}$ is a quadratic objective function for the $\vec{m}_k^\mu$'s, with nonlinear constraints (13) and (14).

Alternative Estimation Approaches

We would like to be able to avoid a nonlinear optimization problem. The cost of this will be to obtain a noiser estimate of the F-matrix than the one provided by the previous gold standard approach.

An initial simplification is to ignore the noise in $\vec{x}_k^L$ for the purpose of estimating the epipolar line $e(\vec{m}_k^L)$. That is, we say the corresponding right image 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

$$(\vec{n}^{\,T}, c)\vec{x}^{\,R} = 0, \tag{15}$$

where

$$\begin{pmatrix} \vec{n} \\ c \end{pmatrix} = \frac{1}{||(I_2\ \vec{0})F^T\vec{x}_k^{\,L}||_2} F^T\vec{x}_k^{\,L}. \tag{16}$$

The normalization in (16) simply ensures $\vec{n}$ is the *unit normal* to the epipolar line $e(\vec{x}_k^{\,L})$. Therefore

$$d(\vec{x}_k^{\,R}, e(\vec{x}_k^{\,L})) \ \equiv\ (\vec{n}^{\,T}, c)\vec{x}^{\,R}, \tag{17}$$

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

We could try to minimize the sum of squares of these epipolar distances $d(\vec{x}_k^{\,R}, e(\vec{x}_k^{\,L}))$ for $k = 1, \ldots, K$. However, due to the normalization factor in (16), the objective function is not quadratic in the unknown F.

Algebraic Error

Consider the reweighted epipolar distance objective function

$$\begin{aligned} \mathcal{O}(F) &\equiv \sum_{k=1}^{K} w(\vec{x}_k^L) d^2(\vec{x}_k^R, e(\vec{x}_k^L)) \\ &= \sum_{k=1}^{K} \left[(\vec{x}_k^L)^T F \vec{x}_k^R \right]^2 . \end{aligned} \tag{18}$$

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

$$w(\vec{x}_k^L) = ||(I_2 \; \vec{0}) F^T \vec{x}_k^L||_2^2. \tag{19}$$

This objective function corresponds to the *algebraic error* in the noiseless epipolar constraint (9).

In terms of maximum likelihood estimation, Equation (18) is appropriate when the variances of the error in algebraic constraints (9) are roughly constant (and the means are zero). If the variances deviate significantly from this, then 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

$$M^\mu = \begin{pmatrix} s^\mu & 0 & b_1^\mu \\ 0 & s^\mu & b_2^\mu \\ 0 & 0 & 1 \end{pmatrix}, \tag{20}$$

   with

$$\frac{1}{K}\sum_{k=1}^{K} M^\mu \vec{x}_k^\mu = (0,0,1)^T, \tag{21}$$

$$\frac{1}{K}\sum_{k=1}^{K}[M^\mu \vec{x}_k^\mu - (0,0,1)^T]_*^2 = (\sigma_1^2, \sigma_2^2, 0)^T, \tag{22}$$

   where $\sigma_1^2 + \sigma_2^2 = 2$. Here $[...]_*^2$ denotes the square of each element. Rescale the image points using $\vec{r}_k^\mu = M^\mu \vec{x}_k^\mu$ for $k = 1, \ldots, K$ and $\mu = L, R$.

2. Minimize the objective function $\mathcal{O}(\hat{F})$

$$\mathcal{O}(\hat{F}) \equiv \sum_{k=1}^{K} \left[(\vec{r}_k^L)^T \hat{F} \vec{r}_k^R\right]^2. \tag{23}$$

   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,

$$F = (M^L)^T \hat{F} M^R \tag{24}$$

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, Multiple View Geometry in Computer Vision, Camb. Univ. Press., 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 to do.

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 (16) and (17) for $1 \leq k \leq K$.
4. Identify inliers

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

5. If the number of inliers $|\text{In}|$ 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 $k \in \text{In}$ (i.e., all inliers). Re-solve for the inlier set In as done in steps 3 and 4 above.

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 $K >> d$ (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

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

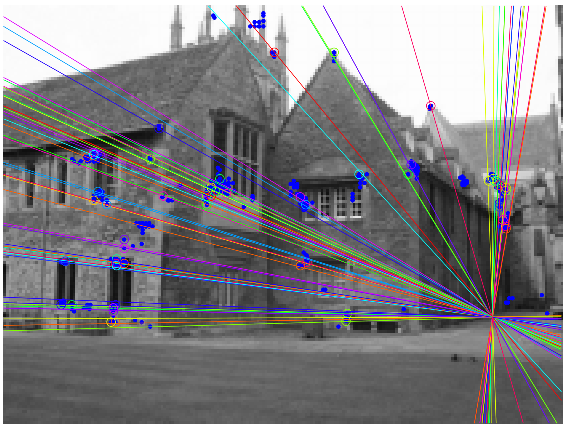
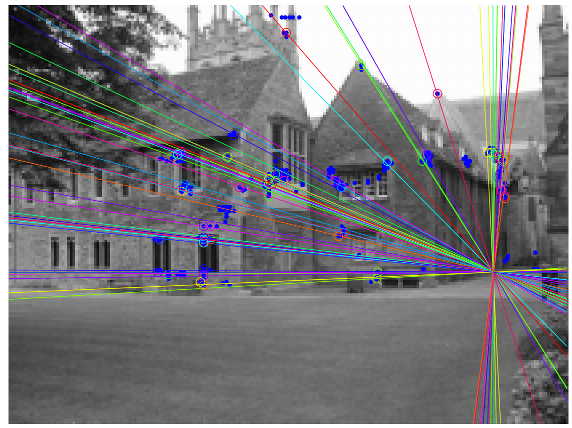
Given an estimate for the fraction of inliers p in the data set, we could then choose T such that $P_0 > 0.95$, say. That is,

$$T > \log(1 - P_0) / \log(1 - p^d).$$

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.

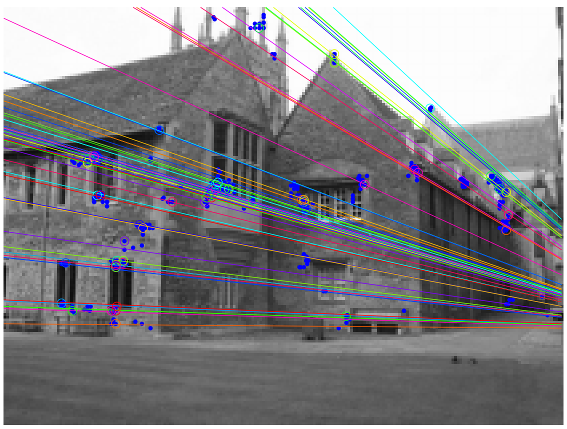
Here have choosen random colours to circle image features. The same colour is then used for the corresponding point in the other image, and also for the epipolar lines generated from these two points.

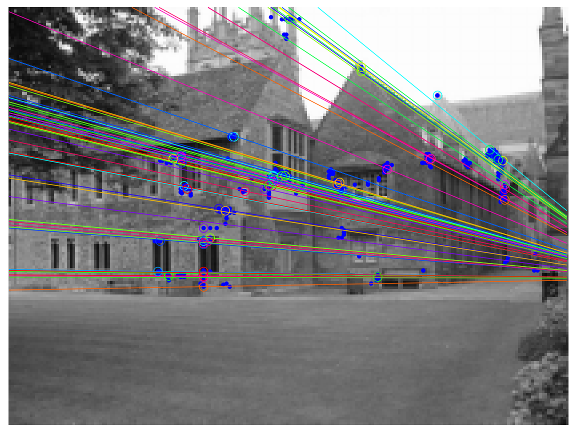
Note:

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.