Computer Vision CITS4240

School of Computer Science & Software Engineering
The University of Western Australia

Self-calibration and the fundamental matrix

In 1992, Olivier Faugeras [1] published a paper that overturned existing thinking about camera calibration and the extraction of metric information from our environment using cameras. He set out to determine what information could be extracted from a binocular stereo rig for which there was no three-dimensional metric calibration data available. All that is assumed is that we have a stereo camera system that is capable, by comparing the two images, of establishing some correspondence between them. Each such correspondence, written (m, m'), indicates that the two image points m and m' are very likely to be the images of the *same* world point M. Thus, the system does *not* know its intrinsic and extrinsic parameters. This is known as the *uncalibrated* system.

Surprisingly, it is still possible to reconstruct some very rich non-metric representations of the environment. What is actually extracted are the *projective invariants* of the scene. Precisely what this means will become clearer as the lecture progresses, but it does indicate that researchers may have been overly optimistic in trying to extract complete metric information; certainly, it has proved to be very difficult and sensitive to noise, and not at all necessary for many applications, such as robot navigation.

It turns out that it is actually possible to use these projective invariants to work out the camera calibration. *Self-calibration* refers to the process of calculating all the intrinsic parameters of the camera using only the information available in the images taken by that camera. No calibration frame or known object is needed: the only requirement is that there is a static object in the scene, and the camera moves around taking images. Thus self-calibration is ideal for a mobile camera, such as a camera mounted on a mobile robot. The actual camera movement itself does not need to be known.

The geometric information that relates two different viewpoints of the same scene is entirely contained in a mathematical construct known as the *fundamental matrix*. The two viewpoints could be a stereo pair of images, or a temporal pair of images. In the latter case the two images are taken at different times with the camera moving between image acquisitions.

We will begin this lecture by considering the geometry of two different images of the same scene, known as the *epipolar geometry*. We will then discuss the fundamental matrix and how it is calculated, and clarify precisely what information can be extracted in the uncalibrated case. Finally, we will consider self-calibration.

Epipolar geometry

Firstly, a word on notation. Points, as entities in their own right, will be denoted in italics. When such points are expressed in Euclidean or projective coordinates, we will use bold notation. Thus a point M in three space might be imaged at m, and m might have coordinates $\mathbf{m} = (u, v)^\top$ or $\mathbf{m} = (u, v, 1)^\top$. Note that projective coordinates are defined up to a scale. So $(u, v, 1)^\top$ and $(su, sv, s)^\top$, for any non-zero scalar s, represent the same image coordinates.

Moreover, as with the last lecture, much of the development in this lecture is done in the setting of projective geometry, which was first introduced in Lecture 1. There is one result which we will use constantly, so it is important to have it clearly understood.

Result: A line going through two points, $\tilde{\mathbf{m}}_1$ and $\tilde{\mathbf{m}}_2$ is represented by the cross product $\tilde{\mathbf{m}}_1 \wedge \tilde{\mathbf{m}}_2$.

Proof. A point on the line is given by $\tilde{\mathbf{x}} = \alpha\tilde{\mathbf{m}}_1 + \beta\tilde{\mathbf{m}}_2$, for arbitrary values of the scalars α and β. This is equivalent to writing that the determinant $|\tilde{\mathbf{x}}, \tilde{\mathbf{m}}_1, \tilde{\mathbf{m}}_2| = 0$. But this determinant can also be written as

$$\tilde{\mathbf{x}}^\top(\tilde{\mathbf{m}}_1 \wedge \tilde{\mathbf{m}}_2) = 0,$$

so the result follows.

In the last lecture, we considered in detail the geometry of a single camera. We will now introduce a second view and study the geometric properties of the set of two views. The main new geometric property is known in computer vision as the *epipolar constraint*.

There are two ways of extracting three-dimensional structure from a pair of images. In the first, and classical method, known as the *calibrated route*, we firstly need to calibrate both cameras (or viewpoints) with respect to some world coordinate system, and from this compute the three-dimensional Euclidean structure of the imaged scene.

However it is the second, or *uncalibrated route*, that more likely corresponds to the way in which biological systems determine three-dimensional structure from vision. In an uncalibrated system, a quantity known as the *fundamental* matrix is calculated from image correspondences, and this is then used to determine the projective three-dimensional structure of the imaged scene.

In both approaches the underlying principle of binocular vision is that of *triangulation*. Given a single image, the three-dimensional location of any visible object point must lie on the straight line that passes through the centre of projection and the image of the object point (see Figure 1). The determination of the intersection of two such lines generated from two independent images is called triangulation.

Clearly, the determination of the scene position of an object point through triangulation depends upon matching the image location of the object point in one image to the location of the same object point in the other image. The process of establishing such matches

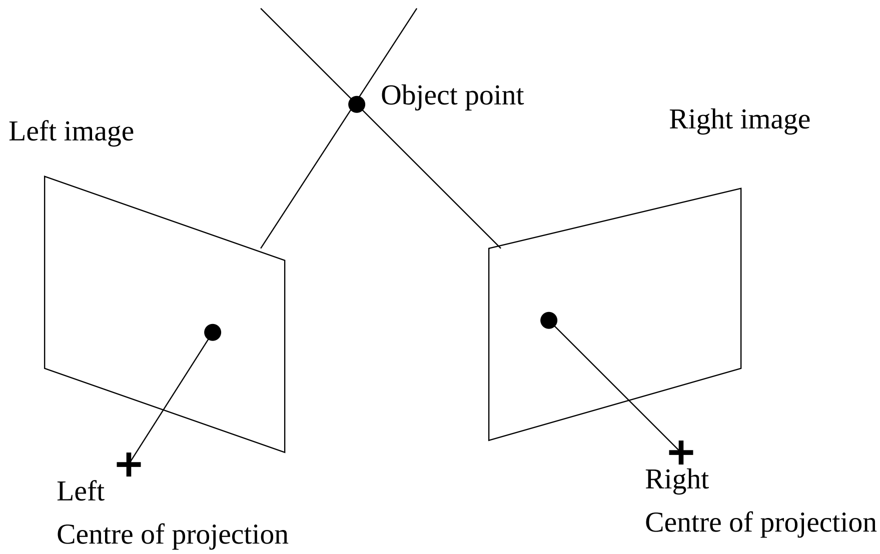

Figure 1: The principle of triangulation in stereo imaging.

between points in a pair of images is called *correspondence*, and is dealt with at length in a different lecture.

At first it might seem that correspondence requires a search through the whole image, but the *epipolar constraint* reduces this search to a single line. To see this, we consider Figure 2.

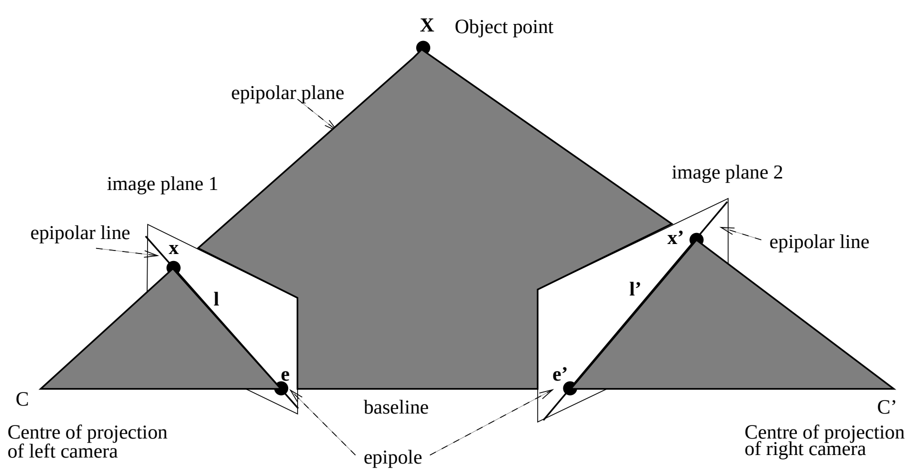

Figure 2: The epipolar constraint.

The *epipole* is the point of intersection of the line joining the optical centres, that is the *baseline*, with the image plane. Thus the epipole is the image, in one camera, of the optical centre of the other camera.

The *epipolar plane* is the plane defined by a 3D point M and the optical centres C and C'. Different 3D points give rise to different epipolar planes. All epipolar planes intersect at the baseline of the binocular system.

The *epipolar line* is the straight line of intersection of the epipolar plane with the image plane. It is the image in one camera of a ray through the optical centre and image point in the other camera. All epipolar lines intersect at the epipole.

Thus, a point $\mathbf{x}$ in one image generates a *line* in the other on which its corresponding point $\mathbf{x}'$ must lie. We see that the search for correspondences is thus reduced from a region to a line. This is illustrated in Figure 3.

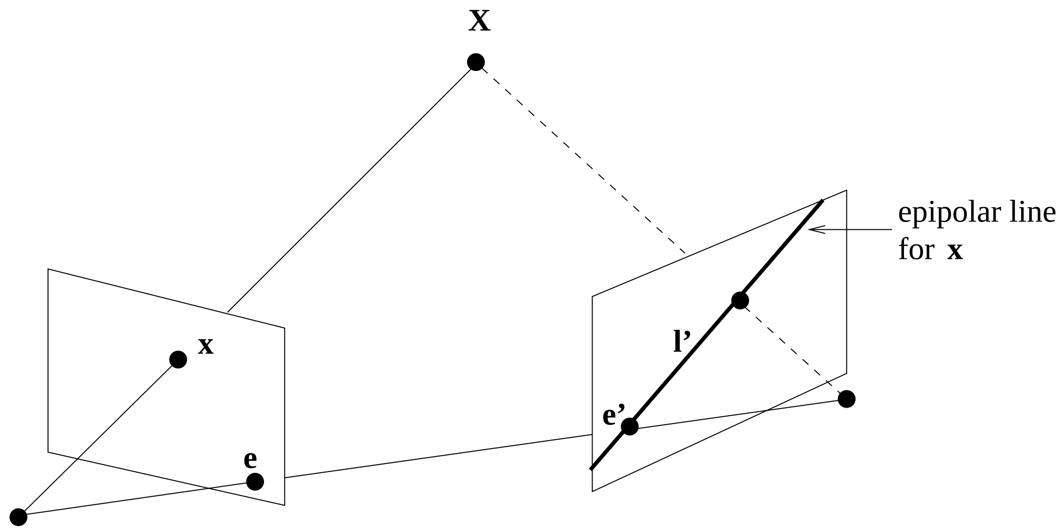

Figure 3: The epipolar line along which the corresponding point for $\mathbf{x}$ must lie.

The essential matrix and the fundamental matrix

To calculate depth information from a pair of images we need to compute the epipolar geometry that embodies the perspective projection of the cameras and the relative orientation and translation between them. If the cameras are calibrated, we capture this geometric constraint in an algebraic representation known as the *essential matrix*. In the uncalibrated environment, it is captured in the *fundamental matrix*.

The essential matrix

Relative to the camera coordinate system fixed at the centre of projection, $\mathbf{C}'$, of the right camera, a scene point M having coordinates $\mathbf{X} = (X, Y, Z, 1)^\top$ is projected onto the image plane via (see the lecture on Camera Calibration)

$$
\begin{aligned}
s' \begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix} &= \begin{bmatrix} f' & 0 & u'_0 & 0 \\ 0 & f' & v'_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} f' & 0 & u'_0 \\ 0 & f' & v'_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} I & \mathbf{0} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \\
\Rightarrow \; s' \begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix} &= K' \begin{bmatrix} I & \mathbf{0} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix},
\end{aligned} \tag{1}
$$

where I denotes the 3×3 identity matrix; f' and (u'_0, v'_0) denote the focal length and principal point of the camera; K' is the camera matrix embodying all the intrinsic parameters of the right camera. Rearranging the equation to eliminate the scalar s' gives

$$
\begin{aligned}
\frac{X}{Z} &= \frac{u' - u'_0}{f'} \\
\frac{Y}{Z} &= \frac{v' - v'_0}{f'}.
\end{aligned} \tag{2}
$$

Thus, relative to the camera coordinate system fixed at $\mathbf{C}'$, the image point m' of M has projective coordinates $\tilde{\mathbf{x}}' = (\frac{u'-u'_0}{f'}, \frac{v'-v'_0}{f'}, 1)^\top = (u' - u'_0, v' - v'_0, f')^\top$.

Similarly, with a camera coordinate system fixed at $\mathbf{C}$, image point m in projective coordinates would be the 3-vector $\tilde{\mathbf{x}} = (u - u_0, v - v_0, f)^\top$.

Suppose that the global reference frame is fixed at the centre of projection $\mathbf{C}'$ of the right camera, then between the two camera coordinate systems is rotation R and translation $\mathbf{t}$ (see Figure 4). That is,

$$
\begin{aligned}
\begin{bmatrix} u' - u'_0 \\ v' - v'_0 \\ f' \end{bmatrix} &= R \begin{bmatrix} u - u_0 \\ v - v_0 \\ f \end{bmatrix} + \mathbf{t} \\
\text{i.e.,} \quad \tilde{\mathbf{x}}' &= R\tilde{\mathbf{x}} + \mathbf{t}.
\end{aligned} \tag{3}
$$

From the proof on Page 2, we see that this is equivalent to $|\tilde{\mathbf{x}}', R\tilde{\mathbf{x}}, \mathbf{t}| = 0$. In other words,

$$
\tilde{\mathbf{x}}'^\top (\mathbf{t} \wedge R\tilde{\mathbf{x}}) = 0, \tag{4}
$$

Note that the cross product of a vector $\mathbf{p} = (p_x, p_y, p_z)^\top$ and a vector $\mathbf{q} = (q_x, q_y, q_z)^\top$ can be expressed as a matrix-vector multiplication:

$$
\mathbf{p} \wedge \mathbf{q} = \begin{bmatrix} 0 & -p_z & p_y \\ p_z & 0 & -p_x \\ -p_y & p_x & 0 \end{bmatrix} \mathbf{q},
$$

where the matrix on the right hand side is the skew-symmetric matrix formed by $\mathbf{p}$. So Equation (4) can be written as

$$
\begin{aligned}
\tilde{\mathbf{x}}'^\top T R \tilde{\mathbf{x}} &= 0 \\
\Rightarrow \; \tilde{\mathbf{x}}'^\top (TR) \tilde{\mathbf{x}} &= 0 \\
\Rightarrow \; \tilde{\mathbf{x}}'^\top \mathbf{E} \tilde{\mathbf{x}} &= 0
\end{aligned} \tag{5}
$$

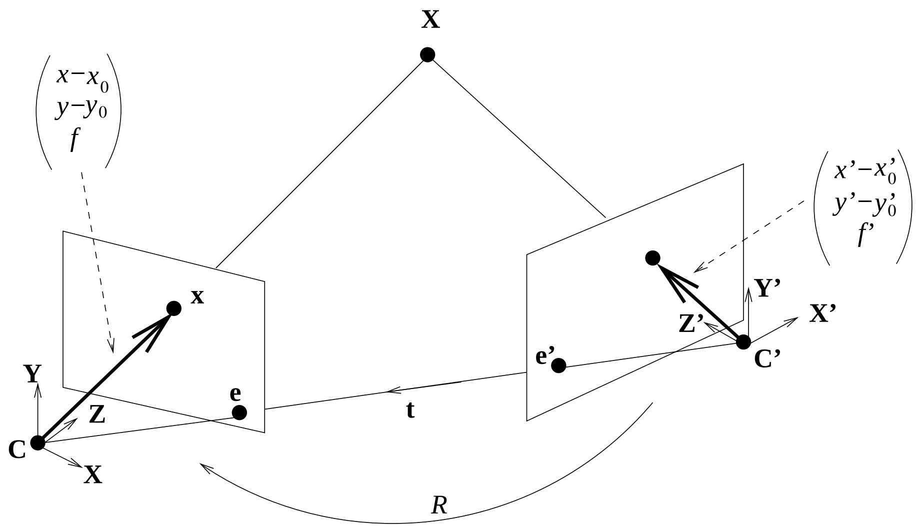

Figure 4: The Euclidean relationship between the two view-centred coordinate systems.

where T is the skew-symmetric matrix formed by $\mathbf{t}$ and $\mathbf{E}$ is a 3×3 matrix known as the *essential* matrix.

Equation (5) is the algebraic representation of epipolar geometry for known calibration, and the essential matrix, which embodies the relative geometry of the two cameras (i.e., the rotation R and translation vector $\mathbf{t}$), relates corresponding image points expressed in the camera coordinate system.

Notice that Equation (5) is homogeneous with respect to $\mathbf{t}$. This reflects the fact that scale is undetermined and we cannot recover the absolute scale of the scene without some extra information, such as knowing the distance in space between two points. Thus $\mathbf{E}$ only depends on five parameters, of which three parameters are from the rotation and two[1] are from the translation. Note also that the case $\mathbf{t} = \mathbf{0}$ is a trivial solution but one from which we are unable to calculate any information about the depth of points in space; for this reason, it is usually excluded.

If the cameras are partially calibrated, i.e., the intrinsic parameters of the cameras are known but the relative geometry between them is not, then we do not have knowledge about R and $\mathbf{t}$; all we have are image coordinates in the image plane. Even so, the essential matrix $\mathbf{E}$ can be estimated from a small number of corresponding points in a similar way as the estimation of the fundamental matrix, as we shall see later on Page 9.

The fundamental matrix

We note that the image point coordinates $(u - u_0, v - v_0, f)^\top$ and $(u' - u'_0, v' - v'_0, f')^\top$ both include the intrinsic parameters of the cameras. In the uncalibrated case, these

[1] Two, rather than three, parameters from the translation $\mathbf{t}$ because of the indeterminable absolute magnitude of $\mathbf{t}$.

parameters are not known. We observe that

$$\begin{bmatrix} u' - u'_0 \\ v' - v'_0 \\ f' \end{bmatrix} = \begin{bmatrix} 1 & 0 & -u'_0 \\ 0 & 1 & -v'_0 \\ 0 & 0 & f' \end{bmatrix} \begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix}$$

Define $\mathbf{x}' = (u', v', 1)^\top$ and since projective coordinates are defined only up to a scale, we have

$$\begin{aligned} \begin{bmatrix} u' - u'_0 \\ v' - v'_0 \\ f' \end{bmatrix} & \sim \frac{1}{f'} \begin{bmatrix} 1 & 0 & -u'_0 \\ 0 & 1 & -v'_0 \\ 0 & 0 & f' \end{bmatrix} \begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix} \\ \Rightarrow \qquad \tilde{\mathbf{x}}' & \sim \begin{bmatrix} \frac{1}{f'} & 0 & -\frac{u'_0}{f'} \\ 0 & \frac{1}{f'} & -\frac{v'_0}{f'} \\ 0 & 0 & 1 \end{bmatrix} \mathbf{x}' \\ \Rightarrow \qquad \tilde{\mathbf{x}}' & \sim K'^{-1}\mathbf{x}', \end{aligned} \tag{6}$$

where $\sim$ denotes equality up to a scale and K' is the camera matrix of the right camera as defined in Equation (1).

Similarly, let $\mathbf{x} = (u, v, 1)^\top$ and let K be the camera matrix of the left camera. Then

$$\tilde{\mathbf{x}} \sim K^{-1}\mathbf{x}. \tag{7}$$

Substituting (7) and (6) into (5) gives

$$\mathbf{x}'^\top K'^{-\top} T R K^{-1} \mathbf{x} = 0.$$

Let

$$\mathbf{F} = K'^{-\top} T R K^{-1}. \tag{8}$$

Then we have

$$\mathbf{x}'^\top \mathbf{F} \mathbf{x} = 0. \tag{9}$$

This 3×3 matrix $\mathbf{F}$ is known as the *fundamental matrix*. Equation (9) is referred to as the *epipolar constraint*.

Note the difference between Equations (5) and (9). Both relate a 3×3 matrix with coordinates of corresponding image points. In (5), the image coordinates $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{x}}'$ include offsets and normalisation by the principal points and focal lengths; in (9), the image coordinates $\mathbf{x}$ and $\mathbf{x}'$ are relative to any arbitrary local image coordinate systems.

It is easy to see the following relationship between the essential matrix and the fundamental matrix:

$$\mathbf{F} \sim K'^{-\top} \mathbf{E} K^{-1}. \tag{10}$$

The essential and the fundamental matrices have the following properties:

(i) The fundamental matrix encapsulates both the intrinsic and the extrinsic parameters of the camera, whilst the essential matrix encapsulates only the extrinsic parameters.

(ii) Both $\mathbf{F}$ and $\mathbf{E}$ are rank-2 matrices as $\text{rank}(T) = 2$. Thus, $\det(\mathbf{F}) = 0$ and $\det(\mathbf{E}) = 0$.

(iii) The essential matrix $\mathbf{E}$ is a 3×3 matrix with only 5 degrees of freedom. To estimate it using corresponding image points, the intrinsic parameters of both cameras must be known.

(iv) The fundamental matrix $\mathbf{F}$ has 7 degrees of freedom. There are 9 matrix elements, but only their ratio is significant, which leaves 8 degrees of freedom. In addition, the constraint that $\det(\mathbf{F}) = 0$ leaves only 7.

(v) $\mathbf{F}$ maps image points to their corresponding epipolar lines, that is, $\mathbf{Fm} = \mathbf{l}'$, since $\mathbf{m}'^{\top}\mathbf{l}' = \mathbf{m}'^{\top}\mathbf{Fm} = 0$. Likewise, $\mathbf{F}^{\top}\mathbf{m}' = \mathbf{l}$ since $\mathbf{l}^{\top}\mathbf{m} = 0$.

(vi) Geometrically, $\mathbf{F}$ maps epipoles to the origin of the corresponding image plane. Algebraically, the epipoles $\mathbf{e}$ and $\mathbf{e}'$ are the null vectors of $\mathbf{F}$ and $\mathbf{F}^{\top}$ respectively. To see this, we note that the epipole $\mathbf{e}$ is the intersection of all the epipolar lines in the left image (see Figure 2). This means that for *any* image coordinates $\mathbf{m}'$ in the right image, $\mathbf{m}'^{\top}\mathbf{Fe} = 0$. This is only possible if $\mathbf{Fe} = \mathbf{0}$. Thus, $\mathbf{e}$ is the null vector of $\mathbf{F}$. Similar argument follows for $\mathbf{e}'$ being the null vector of $\mathbf{F}^{\top}$.

We can also derive the relationship between the fundamental matrix and the two 3×4 projective matrices $\tilde{\mathbf{P}}$ and $\tilde{\mathbf{P}}'$ that we encountered earlier in the Camera Calibration lecture.

Suppose we have two views of a point M in three dimensional space, with M imaged at m in view 1 and m' in view 2. From the last lecture we know that there are projective matrices $\tilde{\mathbf{P}} = [\,\mathbf{P}\;\;\mathbf{p}\,]$ and $\tilde{\mathbf{P}}' = [\,\mathbf{P}'\;\;\mathbf{p}'\,]$ such that in projective coordinates

$$(u, v, 1)^{\top} \sim \mathbf{m} = \tilde{\mathbf{P}}\mathbf{M}$$

and

$$(u', v', 1)^{\top} \sim \mathbf{m}' = \tilde{\mathbf{P}}'\mathbf{M}.$$

Also, the coordinates, $\tilde{\mathbf{C}} = (\mathbf{C}, 1)^{\top} = (C_x, C_y, C_z, 1)^{\top}$ and $\tilde{\mathbf{C}}' = (\mathbf{C}', 1)^{\top} = (C'_x, C'_y, C'_z, 1)^{\top}$, of the two optical centres C and C' are obtained, in the world reference frame, by solving the two systems of linear equations

$$\tilde{\mathbf{P}}\tilde{\mathbf{C}} = \mathbf{0},$$

and

$$\tilde{\mathbf{P}}'\tilde{\mathbf{C}}' = \mathbf{0}.$$

Thus, since $\tilde{\mathbf{P}}\begin{bmatrix} \mathbf{C} \\ 1 \end{bmatrix} = 0$, we can rewrite this as $\mathbf{C} = -\mathbf{P}^{-1}\mathbf{p}$.

So, given a point m in the first view, its corresponding epipolar line l' can be computed from two points known to lie on it:

- One of these is the epipole $\mathbf{e}'$ and is given by

$$\mathbf{e}' = \tilde{\mathbf{P}}'\begin{bmatrix} \mathbf{C} \\ 1 \end{bmatrix} = \tilde{\mathbf{P}}'\begin{bmatrix} -\mathbf{P}^{-1}\mathbf{p} \\ 1 \end{bmatrix}.$$

- Another is the projection onto view 2 of the point at infinity, $\tilde{\mathbf{M}}_\infty = (\mathbf{M}_\infty, 0)^\top$, of the optical ray joining the optical centre C and the point m. Being a point at infinity, the last (4^{th}) component of $\mathbf{M}_\infty$ is 0. Thus,

$$\begin{aligned} \mathbf{m} &= \tilde{\mathbf{P}}\tilde{\mathbf{M}}_\infty = \mathbf{P}\mathbf{M}_\infty \\ \Rightarrow \mathbf{M}_\infty &= \mathbf{P}^{-1}\mathbf{m} \end{aligned}$$

The image coordinates, $\mathbf{m'}_\infty$, of this point in the second image plane is given by

$$\mathbf{m'}_\infty = \mathbf{P'}\mathbf{M}_\infty = \mathbf{P'}\mathbf{P}^{-1}\mathbf{m}. \tag{11}$$

The coordinates of the corresponding epipolar line l' are then given by the cross product of these two points. So we have

$$\mathbf{l'} = \mathbf{e'} \wedge \mathbf{m'}_\infty.$$

By substituting (11) into the above equation and using the fact that $\mathbf{l'} = \mathbf{Fm}$ from property (v), we have

$$\begin{aligned} \mathbf{Fm} &= \mathbf{l'} = \mathbf{e'} \wedge (\mathbf{P'}\mathbf{P}^{-1}\mathbf{m}) \\ \Rightarrow \mathbf{F} &= \begin{bmatrix} 0 & -e'_z & e'_y \\ e'_z & 0 & -e'_x \\ -e'_y & e'_x & 0 \end{bmatrix} \mathbf{P'}\mathbf{P}^{-1}. \end{aligned} \tag{12}$$

Calculating the fundamental matrix

Equation (9) can be written as

$$\mathbf{a}^\top \mathbf{f} = 0,$$

where

$$\begin{aligned} \mathbf{a} &= [uu', vu', u', uv', vv', v', u, v, 1]^\top \text{ and} \\ \mathbf{f} &= [F_{11}, F_{12}, F_{13}, F_{21}, F_{22}, F_{23}, F_{31}, F_{32}, F_{33}]^\top. \end{aligned}$$

This equation is linear and homogeneous in the 9 unknown coefficients of the matrix $\mathbf{F}$. Thus, if we are given at least 8 matches we will, in general, be able to determine a unique solution to $\mathbf{F}$, defined up to a scale factor, by finding the null vector of the data matrix A as defined below:

$$A\,\mathbf{f} = \begin{bmatrix} u_1u'_1 & v_1u'_1 & u'_1 & u_1v'_1 & v_1v'_1 & v'_1 & u_1 & v_1 & 1 \\ u_2u'_2 & v_2u'_2 & u'_2 & u_2v'_2 & v_2v'_2 & v'_2 & u_2 & v_2 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u_nu'_n & v_nu'_n & u'_n & u_nv'_n & v_nv'_n & v'_n & u_n & v_n & 1 \end{bmatrix} \mathbf{f} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{13}$$

Since the fundamental matrix has 7 degrees of freedom, a minimum number of 7 matches can be used to compute $\mathbf{F}$. However, this would require incorporating the constraint $\det(\mathbf{F}) = 0$, which is a cubic polynomial of the elements of $\mathbf{F}$. An easier approach would therefore be employing the linear method described above and then imposing the condition that $\det(\mathbf{F}) = 0$ as a post-process. A final nonlinear method that iteratively refines the estimate of $\mathbf{F}$ while minimizing the reprojection errors is also recommended.

The essential matrix $\mathbf{E}$ can be similarly estimated, i.e., firstly using the linear method, then imposing the $\det(\mathbf{E}) = 0$ constraint, and a final nonlinear optimization. In addition, the final estimated $\mathbf{E}$ must conform to the product of a skew-symmetric matrix and a rotation matrix.

What information can be extracted from two scenes?

In 1996, Faugeras and Robert [4] completely solved the problem about what could be predicted about a third view of an object, given two other views. This problem is central to stereo, motion, and object recognition. They describe the geometry of three cameras, illustrated in Figure 5, as follows: Denoting the cameras by 1, 2 and 3, there are now

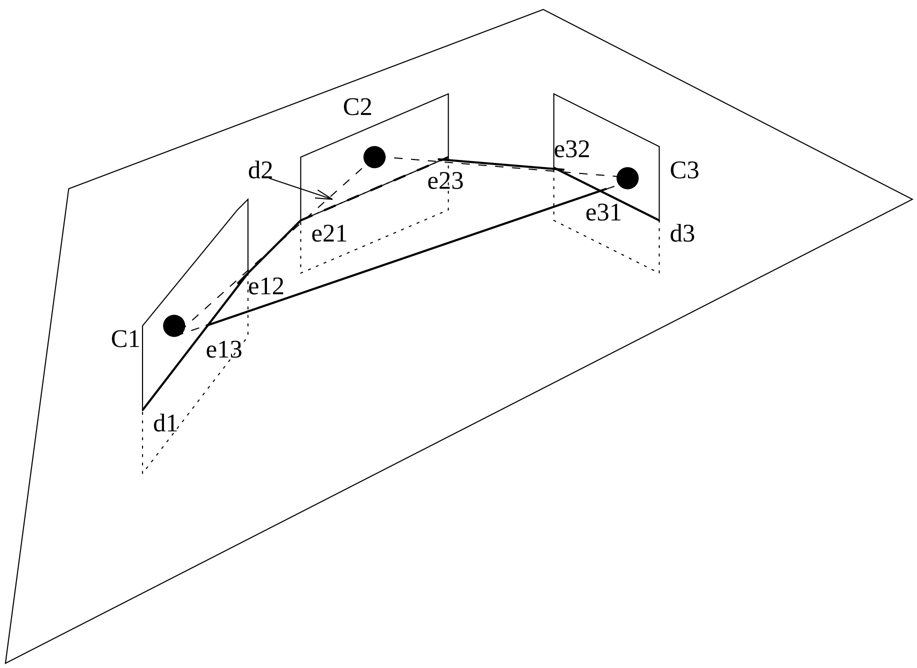

Figure 5: The geometry of three camera systems. There are three optical centres, six epipoles, and three particular epipolar lines.

three fundamental matrices $\mathbf{F}_{ij}$, with the obvious convention on the indices. If m_i is a pixel in image i, its epipolar line in image j is represented by $\mathbf{F}_{ij}\tilde{\mathbf{m}}_i$. Note that we have $\mathbf{F}_{ij} = \mathbf{F}_{ji}^{\top}$. The plane containing the three optical centres is called the *trifocal plane*. It intersects each image plane along a line d_i which contains the epipoles e_{ii+1} and e_{ii+2} of camera i with respect to cameras $i + 1$ and $i + 2$ (all mod 3). Because of the epipolar geometry, we have

$$\mathbf{F}_{ii+1}\tilde{\mathbf{e}}_{ii+2} = \mathbf{d}_{i+1} = \tilde{\mathbf{e}}_{i+1,i} \wedge \tilde{\mathbf{e}}_{i+1,i+2}.$$

In this work it is assumed that the three fundamental matrices are known, but not that the system is fully calibrated. It is shown that it is possible to predict how a scene would look from a third viewpoint, given two other views. They consider predicting the third view of points, lines, and curvatures.

The prediction of points is very simple. We assume that we have two corresponding pixels m_1 and m_2 in images 1 and 2. Them m_3 must belong to the epipolar line of m_1 in the third image, given by $\mathbf{F}_{13}\tilde{\mathbf{m}}_1$, and to the epipolar line of pixel m_2 in the third image, given by $\mathbf{F}_{23}\tilde{\mathbf{m}}_2$. Therefore, m_3 belongs to the intersection of these two lines, and we can write

$$\tilde{\mathbf{m}}_3 = \mathbf{F}_{13}\tilde{\mathbf{m}}_1 \wedge \mathbf{F}_{23}\tilde{\mathbf{m}}_2.$$

The prediction of lines is also simple. We assume now that we are given two corresponding lines l_1 and l_2 in images 1 and 2. The problem is to determine the position of l_3 in image 3. Let m_1, m_1' be two points of l_1. They define two points m_2, m_2' of l_2 as the intersections of the epipolar line of m_1 represented by $\mathbf{F}_{12}\tilde{\mathbf{m}}_1$ and of m_1' represented by $\mathbf{F}_{12}\tilde{\mathbf{m}}_1'$ with l_2. Therefore we can write

$$\tilde{\mathbf{m}}_2 = \mathbf{F}_{12}\tilde{\mathbf{m}}_1 \wedge \mathbf{l}_2$$

and

$$\tilde{\mathbf{m}}_2' = \mathbf{F}_{12}\tilde{\mathbf{m}}_1' \wedge \mathbf{l}_2.$$

Therefore, the line l_3 is defined by the two points m_3 and m_3', intersections of the epipolar lines of m_1 and m_1' and m_2 and m_2' in the third image. Therefore we can write

$$\mathbf{l}_3 = \tilde{\mathbf{m}}_3 \wedge \tilde{\mathbf{m}}_3'.$$

The prediction of curvatures is slightly more complicated and will not be covered in these lectures.

Self-calibration

Maybank and Faugeras [7] showed that self-calibration can be computed from a single uncalibrated camera that undergoes some displacement. All that is assumed is that some point matches can be established between images. They developed an algorithm that requires the camera to undergo a minimum of three displacements; there must be a minimum of seven matched points between successive images. From this information they are able to extract the intrinsic parameters of the camera.

Since the pioneer work of Faugeras [1], much research papers have been reported in the literature on camera self-calibration and cannot be all covered in this lecture. More recent texts in this area include [3, 5].

References

[1] Olivier Faugeras. *Three Dimensional Computer Vision*. MIT Press, 1993.

[2] Olivier Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? *Procs. ECCV 92, Lecture Notes in Computer Science*, 588, Springer-Verlag, 1992.

[3] Olivier Faugeras and Quang-Tuan Luong. The geometry of multiple images. MIT Press, 2001.

[4] Olivier Faugeras and Luc Robert. What can two images tell us about a third one? *International Journal of Computer Vision*, 18:5-19, 1996.

[5] Richard Hartley and Andrew Zisserman. Multiple view geometry in computer vision. Cambridge University Press, 2nd Ed, 2004.

[6] Quang-Tuan Luong, Rachid Deriche, Olivier Faugeras, and Theodore Papadopoulo. On determining the fundamental matrix: Analysis of different methods and experimental results. INRIA Rapport de recherche, No. 1894, Mai 1993.

[7] Stephen J. Maybank and Olivier D. Faugeras. A theory of self-calibration of a moving camera. *International Journal of Computer Vision*, 8:2, 123-151, 1992.