3.6 Image rectification

3.6.1 Horizontal epipolar lines

In the special case when the image planes $\mathcal{P}_1$ and $\mathcal{P}_2$ are coplanar and parallel to the vector C_1C_2 joining the optical centers, the epipoles occur at infinity and the epipolar lines form a bundle of parallel lines. If, in addition, the image coordinate systems have been judiciously chosen, then the epipolar lines can be made horizontal, so that a point (u'_1, v'_1) of image 1 has the segment $v'_2 = v'_1$ of image 2 as its conjugate epipolar line. We thus have the situation shown in figure 3.6.

In this section, we show that for each image there always exists a linear transformation in projective coordinates producing horizontal conjugate epipolar lines. (F. Lustman [Lus87] showed in his thesis that the rectification could be carried out somewhat differently by explicitly using the intrinsic and extrinsic camera parameters, and other approaches have been proposed by Gaillat, Caprile and Torre [Gai88, CT88].)

Consider figure 3.9, in which the optical centers C_i and the image planes $\mathcal{P}_i$ are represented for each camera. The principle of image rectification is to define new perspective matrices M and N which preserve the two optical centers C_1 and C_2 respectively, but with a new unique image plane $\mathcal{P}'$ parallel to C_1C_2. Rectification is then the operation of transformation from coordinates (u_i, v_i) to new coordinates (u'_i, v'_i) in each image i.

3.6.2 New perspective matrices

The constraints on the new perspective matrices M and N are:

1. The optical centers of M and N are C_1 and C_2 respectively (to give a unique match between image points I_i and I'_i respectively before and after rectification).
2. The focal plane of M is identified with that of N (to produce parallel epipolar lines in both images).
3. For any point P (not in the optical plane), the image points I'_1 and I'_2 obtained by M and N respectively are such that $v'_1 = v'_2$ (to simplify the computation of the epipolar lines as much as possible).

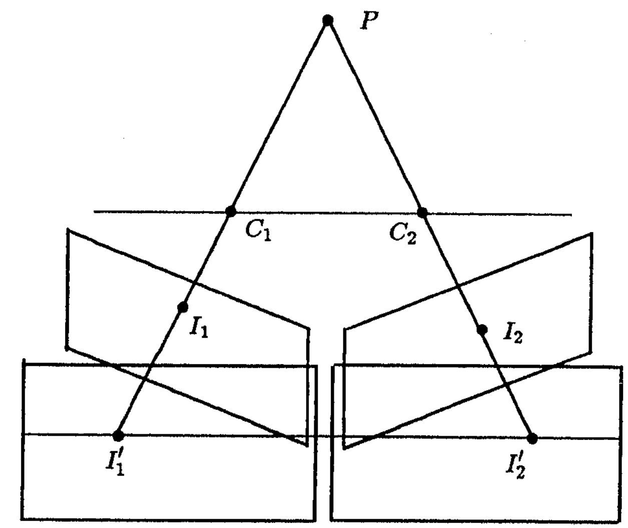

Figure 3.9
Rectification of two images.

To explain the constraints on the matrices M and N, let

$$M = \begin{pmatrix} m_1^t & m_{14} \\ m_2^t & m_{24} \\ m_3^t & m_{34} \end{pmatrix} \qquad N = \begin{pmatrix} n_1^t & n_{14} \\ n_2^t & n_{24} \\ n_3^t & n_{34} \end{pmatrix}$$

There are 24 parameters to evaluate. Now, we have seen that the perspective matrices were defined to within a factor of scale. Thus, for example, we may set[4]

$$m_{34} = n_{34} = \|C_1 \times C_2\|^2 \qquad (3.6.12)$$

There remain 22 parameters to compute in order to define the matrices M and N. Now, constraint 2 implies that

$$\forall P \quad m_3^t P + \|C_1 \times C_2\|^2 = 0 \Longleftrightarrow n_3^t P + \|C_1 \times C_2\|^2 = 0$$

Thus

$$m_3 = n_3 \qquad (3.6.13)$$

[4] Providing that $\|C_1 \times C_2\| \neq 0$.

In addition, constraint 3 implies that

$$\begin{aligned} \forall P \quad m_3^t\,P + \|C_1 \times C_2\|^2 &\neq 0 \\ \frac{m_2^t\,P + m_{24}}{m_3^t\,P + \|C_1 \times C_2\|^2} &= \frac{n_2^t\,P + n_{24}}{n_3^t\,P + \|C_1 \times C_2\|^2} \end{aligned}$$

Now from $m_3 = n_3$, we have

$$m_2 = n_2 \quad \text{and} \quad m_{24} = n_{24} \tag{3.6.14}$$

Finally, constraint 1 can be written:

$$M \begin{pmatrix} C_1 \\ 1 \end{pmatrix} = N \begin{pmatrix} C_2 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Taking equations 3.6.13 and 3.6.14 into account, from constraint 1 we get the six equations:

$$\begin{aligned} m_1^t\,C_1 + m_{14} &= 0 \\ m_2^t\,C_1 + m_{24} &= 0 \\ m_2^t\,C_2 + m_{24} &= 0 \\ m_3^t\,C_1 + \|C_1 \times C_2\|^2 &= 0 \\ m_3^t\,C_2 + \|C_1 \times C_2\|^2 &= 0 \\ n_1^t\,C_2 + n_{14} &= 0 \end{aligned} \tag{3.6.15}$$

In conclusion, equations 3.6.12 express 15 linear equations in the 24 parameters of M and N. Nine degrees of liberty remain, which correspond to the available degrees of liberty for the orientation and the distance of the plane $\mathcal{P}'$, as well as the coordinate systems in the new images.

We will set the available degrees of liberty as simply as possible. For example, let

$$m_{14} = m_{24} = n_{14} = 0$$

In this case, it is clear that the following properties hold:

1. m_1 must be orthogonal to C_1,

2. n_1 must be orthogonal to C_2,
3. m_2 must be orthogonal to C_1 and C_2,
4. m_3 must be orthogonal to $C_1 - C_2$.

To verify property 3, let

$$m_2 = (C_1 \times C_2)$$

To verify property 1, let m_1 be orthogonal to C_1, and to avoid a degenerate perspective matrix, let m_1 be orthogonal to m_2, which gives

$$m_1 = (C_1 \times C_2) \times C_1$$

Applying the same argument to property 2 gives:

$$n_1 = (C_1 \times C_2) \times C_2$$

Next, to verify property 4, let m_3 be the vector product of $C_1 - C_2$ with a vector u such that $m_3^t\,C_1 + \|C_1 \times C_2\|^2 = 0$, which gives $u = C_1 \times C_2$ and thus:

$$m_3 = (C_1 - C_2) \times (C_1 \times C_2)$$

Finally, matrices M and N are defined by

$$M = \begin{pmatrix} ((C_1 \times C_2) \times C_1)^t & 0 \\ (C_1 \times C_2)^t & 0 \\ ((C_1 - C_2) \times (C_1 \times C_2))^t & \|C_1 \times C_2\|^2 \end{pmatrix} \tag{3.6.16}$$

$$N = \begin{pmatrix} ((C_1 \times C_2) \times C_2)^t & 0 \\ (C_1 \times C_2)^t & 0 \\ ((C_1 - C_2) \times (C_1 \times C_2))^t & \|C_1 \times C_2\|^2 \end{pmatrix} \tag{3.6.17}$$

3.6.3 Rectification

As we have already seen in section 3.4.3, an image point $I_1(u_1, v_1)$ of image 1 comes from a point $P(x, y, z)$ with coordinates

$$P = C_1 + \lambda\,n$$

where n, a direction vector of the segment I_1C_1, is given by equation 3.4.4. The projective coordinates of the new image I_1' of point P are then given by

$$I_1'^* = \begin{pmatrix} U_1' \\ V_1' \\ S_1' \end{pmatrix} = M \begin{pmatrix} C_1 + n \\ 1 \end{pmatrix}$$

Now, since C_1 is the optical center of M, we have $M\,(C_1, 1)^t = 0$. As a consequence, the computation of I' reduces to

$$I_1'^* = \begin{pmatrix} U_1' \\ V_1' \\ S_1' \end{pmatrix} = M'\,n$$

where M' is the 3×3 matrix obtained by removing the last column of M. Since n is computed by an affine transformation of the coordinates (u_1, v_1), we note that it is sufficient to put (cf. equation 3.4.4)

$$Q_1 = \begin{pmatrix} ((C_1 \times C_2) \times C_1)^t \\ (C_1 \times C_2)^t \\ ((C_1 - C_2) \times (C_1 \times C_2))^t \end{pmatrix} [t_2^1 \times t_3^1 \quad t_3^1 \times t_1^1 \quad t_1^1 \times t_2^1]$$

a 3×3 matrix, to obtain

$$\begin{pmatrix} U_1' \\ V_1' \\ S_1' \end{pmatrix} = Q_1 \begin{pmatrix} u_1 \\ v_1 \\ 1 \end{pmatrix} \qquad (3.6.18)$$

Proceeding absolutely symmetrically, putting

$$Q_2 = \begin{pmatrix} ((C_1 \times C_2) \times C_2)^t \\ (C_1 \times C_2)^t \\ ((C_1 - C_2) \times (C_1 \times C_2))^t \end{pmatrix} [t_2^2 \times t_3^2 \quad t_3^2 \times t_1^2 \quad t_1^2 \times t_2^2]$$

gives

$$\begin{pmatrix} U_2' \\ V_2' \\ S_2' \end{pmatrix} = Q_2 \begin{pmatrix} u_2 \\ v_2 \\ 1 \end{pmatrix} \qquad (3.6.19)$$

Rectification of images 1 and 2 thus reduces to the application of the two linear transformations in projective coordinates given by the equations 3.6.18 and 3.6.19.

3.6.4 Three cameras

For three cameras, it is possible to rectify the images to have horizontal epipolar lines between images 1 and 2, and vertical epipolar lines between images 1 and 3. It is sufficient to follow the same steps as for two cameras, and to compute the new perspective matrices M, N and Q from the optical centers C_1, C_2 and C_3 as follows:

$$M = \begin{pmatrix} (C_3 \times C_1)^t & 0 \\ (C_1 \times C_2)^t & 0 \\ (C_1 \times C_2 + C_2 \times C_3 + C_3 \times C_1)^t & -(C_1, C_2, C_3) \end{pmatrix} \qquad (3.6.20)$$

$$N = \begin{pmatrix} (C_2 \times C_3)^t & 0 \\ (C_1 \times C_2)^t & 0 \\ (C_1 \times C_2 + C_2 \times C_3 + C_3 \times C_1)^t & -(C_1, C_2, C_3) \end{pmatrix} \qquad (3.6.21)$$

$$Q = \begin{pmatrix} (C_3 \times C_1)^t & 0 \\ (C_2 \times C_3)^t & 0 \\ (C_1 \times C_2 + C_2 \times C_3 + C_3 \times C_1)^t & -(C_1, C_2, C_3) \end{pmatrix} \qquad (3.6.22)$$

In these equations, we denote by (C_1, C_2, C_3) the triple product of vectors C_1, C_2 and C_3.

In equations 3.6.20 to 3.6.22, we chose matrix Q to produce equality between the ordinate v_3' of I_3' and the abscissa u_2' of I_2', i.e., $v_3' = u_2'$. Note that by setting $C_3 = C_1 \times C_2$ in equations 3.6.16 and 3.6.17 we again get the matrices M and N used for rectifying two images.[5]

The rectification itself operates in exactly the same way as previously (cf. equations 3.6.18 and 3.6.19), but with the three 3×3 rectification matrices R_1, R_2, and R_3, written

$$R_i = \begin{pmatrix} (C_{i-1} \times C_i)^t \\ (C_i \times C_{i+1})^t \\ (C_1 \times C_2 + C_2 \times C_3 + C_3 \times C_1)^t \end{pmatrix} [t_2^i \times t_3^i \quad t_3^i \times t_1^i \quad t_1^i \times t_2^i] \qquad (3.6.23)$$

with the conventions that $i + 1 = 1$ if $i = 3$ and $i - 1 = 3$ if $i = 1$.

[5] To within a small difference in their third lines.

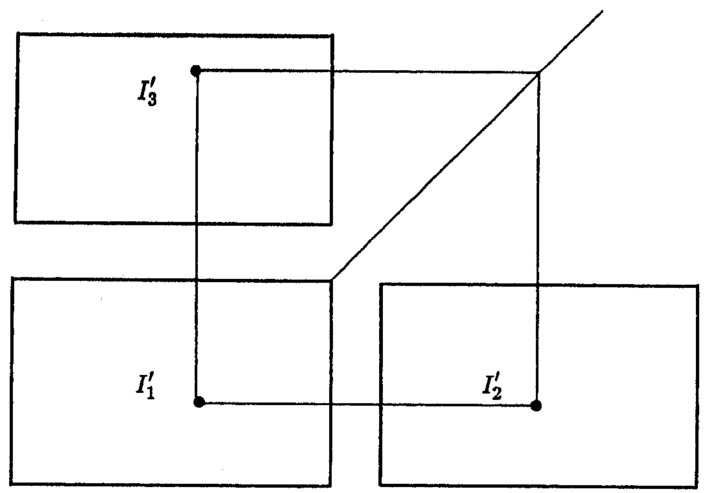

Figure 3.10
After rectification of three images.

Thus for three homologous points I_1', I_2', and I_3', we have the relations:

$$\begin{aligned} v_2' &= v_1' \\ u_3' &= u_1' \\ v_3' &= u_2' \end{aligned} \qquad (3.6.24)$$

and after rectification of the three images, we have the situation shown in figure 3.10.

3.6.5 Algorithmic complexity

The rectification of k images ($k = 2$ or 3) necessitates storing k 3×3 matrices, i.e., $9k$ parameters, and we need 6 multiplications, 6 additions and 2 divisions per rectified point.

3.6.6 Intrinsic rectification and limitations

The M, N and Q used for rectification depend on the choice of the origin O of the absolute coordinate system of the scene. An important question arises: how to free ourselves from this dependence?

One solution consists in choosing a point O', which depends intrinsically on the relative geometry of the cameras, as the new origin of the scene. For example, O' may be the image point which minimizes the sum of the squares of the distances relative to the optical axes of the two (or three) cameras. This idea is developed in [AH88].

Another question is to determine the domain of validity of the proposed rectification. There exist camera positions for which rectification is meaningless (e.g., two cameras sharing a common optical center, i.e., $C_1 = C_2$), or produces very distorted images (e.g., two cameras with optical centers aligned along a common optical axis). Such situations correspond to singularities of the rectification matrices Q_1, Q_2, and R_i (which would exhibit a vanishing third line and could therefore be easily checked prior to rectification).

3.6.7 Example

The following figure shows the contours of a pair of images before and after rectification. The rectification was carried out by applying the transformations in equations 3.6.18 and 3.6.19.

Figure 3.12 shows a triplet of contours before and after rectification. Here, we have applied the transformations in equations 3.6.23.

3.7 Reconstruction in three dimensions

Knowledge of T_1 and T_2 is sufficient to compute the three coordinates of any point P, given its two images I_1 and I_2.

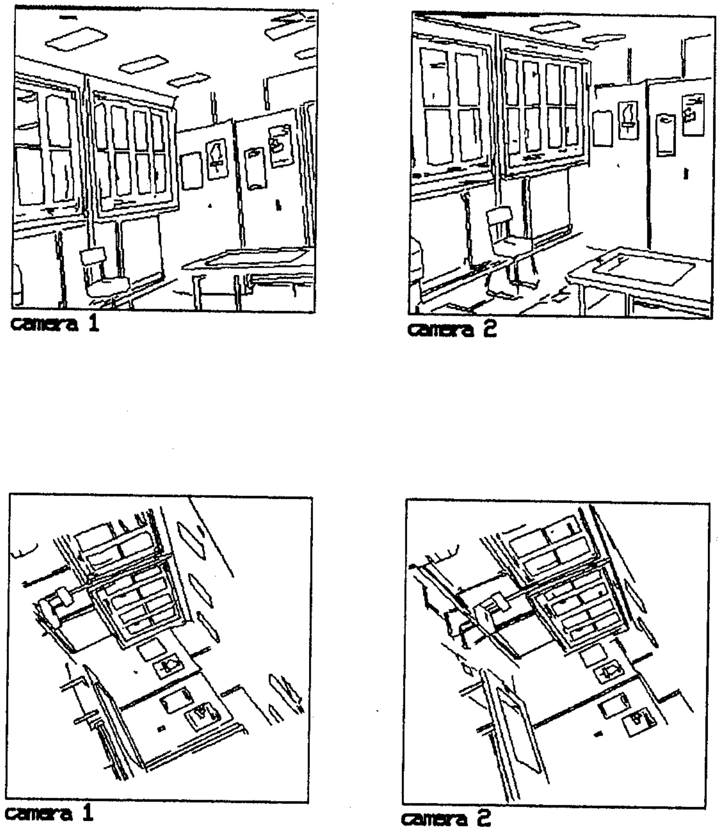

Figure 3.11
Pairs of contours before and after rectification.

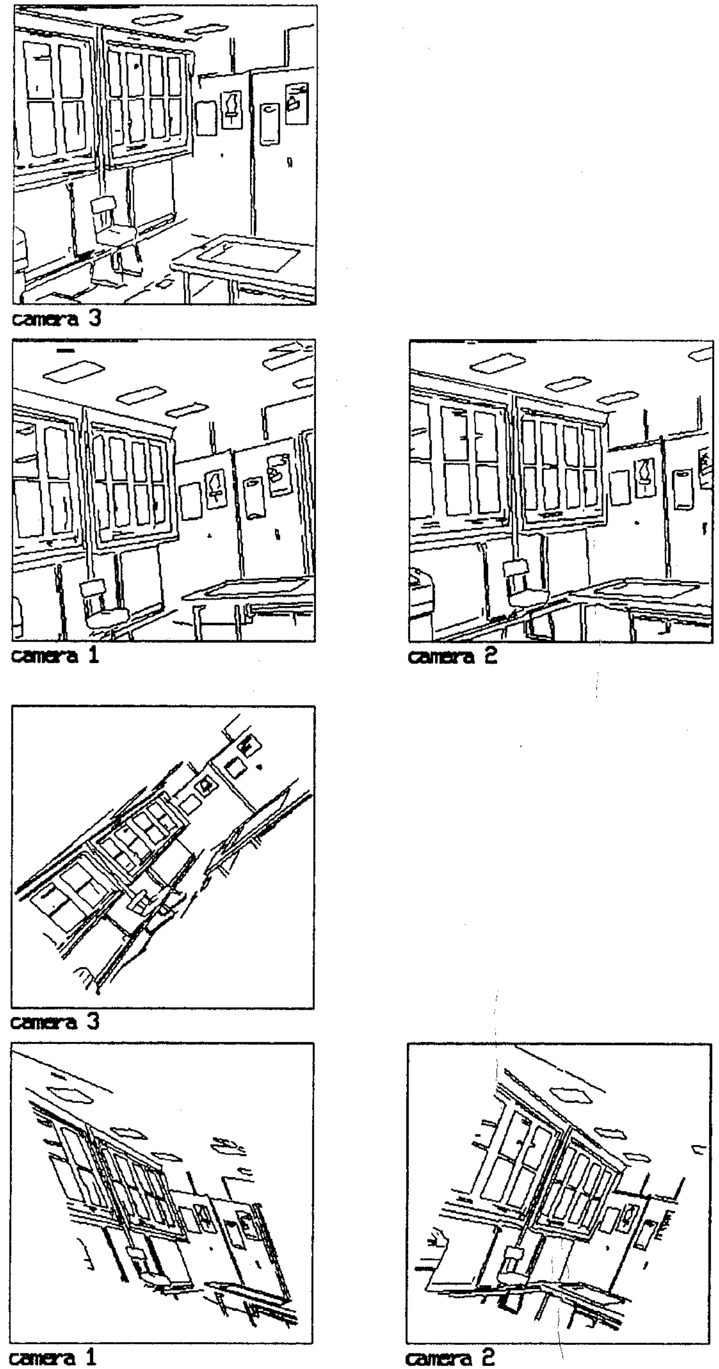

Figure 3.12
Triplets of contours before and after rectification.

In effect, if we write the system 3.3.1 for cameras 1 and 2, we obtain a new system of 4 equations in the three unknown coordinates (x, y, z) of P

$$\begin{aligned}
(t_1^1 - u_1 t_3^1)^t P + t_{14}^1 - u_1 t_{34}^1 &= 0 \\
(t_2^1 - v_1 t_3^1)^t P + t_{24}^1 - v_1 t_{34}^1 &= 0 \\
(t_1^2 - u_2 t_3^2)^t P + t_{14}^2 - u_2 t_{34}^2 &= 0 \\
(t_2^2 - v_2 t_3^2)^t P + t_{24}^2 - v_2 t_{34}^2 &= 0
\end{aligned}$$

in which the index j of t_i^j refers to camera j.

In theory, these equations are related since I_1 and I_2 are chosen in the same epipolar plane. However, numerical imprecision and the absence of an objective criterion in the choice of the equation to eliminate suggests solving the whole system either by least squares or by Kalman filtering. This approach extends naturally to trinocular stereo vision, and more generally to reconstruction based on an arbitrary number of cameras.

3.7.1 Computation by least squares

For n cameras, set

$$Aa = b$$

with $a = (x, y, z)^t$ and

$$A = \begin{pmatrix} A_1 \\ \vdots \\ A_n \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$$

for which

$$A_i = \begin{pmatrix} (t_1^i - u_i t_3^i)^t \\ (t_2^i - v_i t_3^i)^t \end{pmatrix} \quad \text{and} \quad b_i = \begin{pmatrix} u_i t_{34}^i - t_{14}^i \\ v_i t_{34}^i - t_{24}^i \end{pmatrix}$$

The least squares solution is then given by

$$a = (A^t A)^{-1} A^t b$$

provided $A^t A$ is invertible.

3.7.2 Computation by Kalman filtering

Consider a as the state vector, and the measurement equations to be of the form $A_i a = b_i + w_i$ where w_i is the noise in the measurements which takes into account the uncertainty in the positions of the image points I_1 and I_2 as well as the uncertainty in the parameters of the matrices T_1 and T_2. More details are given in the second part of the book.

3.8 Summary

The geometric constraints of stereo vision (i.e., the computation of the epipolar lines) are exploited by using the matrices T_1 and T_2 to compute the coordinates of the epipoles E_1 and E_2, and using the 2×3 matrices M_{12} and M_{21} to compute the direction vectors of the epipolar lines. Thus, we store only 16 numbers, and the parameters of an epipolar segment are computed by only 4 multiplications and 4 additions.

This result is valid in the general case, when the epipoles are defined in the images (S_{E_1} and $S_{E_2} \neq 0$). In the special case where $S_{E_i} = 0$, we need the 3×3 matrix N_{ij} instead of M_{ij}. In this case, we store three additional numbers, and the parameters of an epipolar segment are computed in image i by 6 multiplications, 6 additions and 2 divisions.

Unfortunately, the geometric constraints are not sufficient to determine stereoscopic matches. Additional physical constraints are necessary (as we shall see in following chapters). Some of these are expressed by computing an interval of allowable disparities in each region, and by computing the disparity gradient in each interval as a function of the distance. The disparity gradient may be stored as a table, which typically requires storing 448 numbers (for two images consisting of 16 regions of seven intervals on average, with each interval consisting of two numbers, one disparity, and a disparity gradient).

Rectifying k images ($k = 2, 3$) necessitates storing k 3×3 matrices, i.e., $9k$ parameters. Each rectified points needs 6 multiplications, 6 additions and 2 divisions.

Finally, to compute the spatial position of the physical points whose images have been matched, we need to store k ($k = 2, 3$) matrices T_i