Artificial Vision for Mobile Robots

Stereo Vision and Multisensory Perception

Nicholas Ayache

Translated from the French by
Peter T. Sander

The MIT Press
Cambridge, Massachusetts
London, England

in parallel. Finally, the results of spatial reconstruction of a scene incorporate an additional measurement and are thus more precise. These decisive advantages are confirmed by further experimental results.

- Chapter 9 concludes the first part with a **synthesis** and a discussion, and presents related work.

3 Calibration

3.1 Introduction

In this chapter, we present a method for calibrating a set of cameras. The method consists in first modeling the process of the formation of each image by a linear transformation in projective coordinates, followed by computation of the parameters of the transformation. These are subsequently used to specify geometric stereo constraints and to determine the spatial position of a point from multiple images.

3.2 Image modeling

Each camera is modeled by its optical center C and its image plane $\mathcal{P}$ (cf. Fig. 3.1). This camera model is termed *pinhole*. A point P in the observed space projects onto the camera retina at an image point I. Point I is the intersection of the line PC with the plane $\mathcal{P}$.

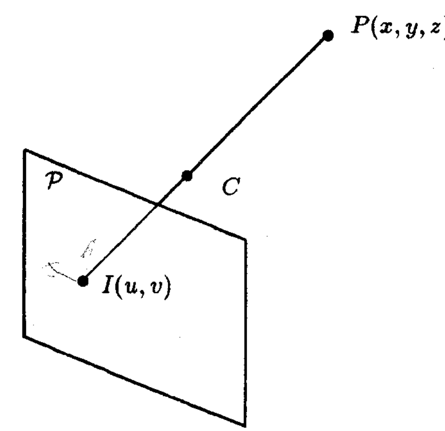

Figure 3.1
Model of image formation.

The transformation from P to I is modeled by a linear transformation T in projective coordinates. Letting $I^* = (U, V, S)^t$ be the projective coordinates of I and $(x, y, z)^t$ be the coordinates of P, we have the relation

$$I^* = \begin{pmatrix} U \\ V \\ S \end{pmatrix} = T \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}$$

where T is a 3×4 matrix, generally called the *perspective matrix* of the camera.

If the line CP is parallel to the image plane $\mathcal{P}$, then $S = 0$ and the coordinates $(u, v)^t$ of the image point I are undefined. In this case, we say that P is in the *focal plane* of the camera. In the general case, $S \neq 0$ and the image coordinates of I are defined by:

$$I = \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} U/S \\ V/S \end{pmatrix}$$

3.3 Determination of T

Calibracros

3.3.1 Principle

T is determined by analyzing the image of a calibration grid where the positions in the image of the intersection points are known with precision. As we see below, this grid must be composed of at least two distinct planes.

The 3×4 matrix T consists of twelve parameters, but these are defined only up to a factor of scale. Thus, an additional constraint must be found. The simplest is to assume that T_{34} is non-zero and to set it to 1 (for a discussion of this constraint and the use of other constraints, see [FT86]). We now have 11 unknown parameters.

Each match of an image point I with an associated scene point P gives two linear equations in the 11 unknowns of the matrix T, i.e.,

$$\begin{aligned} P^t t_1 + t_{14} - u(P^t t_3 + 1) &= 0 \\ P^t t_2 + t_{24} - v(P^t t_3 + 1) &= 0 \end{aligned} \qquad (3.3.1)$$

where t_{ij} is the (i, j) element of T, and t_i is the vector composed of the first three elements of the row i of T:

$$t_i = (t_{i1}, t_{i2}, t_{i3})^t.$$

Thus in theory, six non-coplanar points suffice (cf. [Fau91]) to determine the 11 parameters of T. In practice, dozens of points are available, and the parameters of T are computed either by least squares or by Kalman filtering.

3.3.2 Computation by least squares

Let a be the vector of unknown parameters:

$$a = \begin{pmatrix} t_1 \\ t_{14} \\ t_2 \\ t_{24} \\ t_3 \end{pmatrix}$$

We denote by P_i the n observed points with images I_i. Thus we have the system of equations:

$$Aa = b$$

with

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

$$A_i = \begin{pmatrix} \bar{P}_i^t & 0_{1\times4} & -u_i P_i^t \\ 0_{1\times4} & \bar{P}_i^t & -v_i P_i^t \end{pmatrix} \quad \text{and} \quad b_i = \begin{pmatrix} u_i \\ v_i \end{pmatrix}$$

In the equations, $\bar{P}_i^t = (x_i, y_i, z_i, 1)$ and $0_{1\times4}$ is the 1×4 zero matrix.

The least squares solution is thus given by

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

providing that $A^t A$ is invertible.

3.3.3 Computation by Kalman filtering

The solution of the above system can be calculated by a recursive technique which takes into account an initial estimate and the noise in the measurements (see Part II of the book). In this approach, a is considered as a state vector, and and the measurement equations are of the form $A_i a = b_i + w_i$ where w_i is the noise in the measurement of the image point (u_i, v_i).

3.3.4 Intrinsic and extrinsic parameters

For certain applications, it is necessary to *explicitly* extract a set of *intrinsic* and *extrinsic* camera parameters from the matrix T.

The extrinsic parameters relate the *absolute* 3D coordinate system of the scene with a *relative* 3D camera coordinate system in order to locate the position of the camera in the scene. Six independent parameters are needed to represent the geometric displacement in $\mathbb{R}^3$, three for the rotation and three for the translation.

The intrinsic parameters relate scene points given in the 3D coordinate system *relative* to the camera with the image points given in the frame of the image plane (the pixels). Typically, five independent parameters are available to model this transformation; they consist of the unit lengths along the u and v axes, the angle between these axes, and the position of the origin of the camera frame.

The 3×4 matrix T, defined up to a factor of scale (relating vectors given in projective coordinates), can thus explicitly supply the eleven extrinsic and intrinsic parameters. In the following, we do not need to compute these parameters explicitly and the interested reader is referred to [Tos87, FT86, Tsa86a, Lus87] for a detailed analysis.

3.4 Epipolar lines

In this section, we are interested in the relations between multiple cameras.

3.4.1 Definition

Given point I_1 in image 1, we seek its match I_2 in image 2 (cf. Fig. 3.2). Point I_2 necessarily belongs to a straight line of image 2 determined completely by I_1, and called the *epipolar line* associated with I_1.

In effect, the set of points P whose image corresponds to I_1 is the line C_1I_1 containing point C_1. The image of this line in camera 2 is the epipolar line associated with I_1. The problem is completely symmetric, and if we consider the plane Q defined by I_1, C_1 and C_2, we see that this plane intersects the image planes along two straight lines, the *conjugate epipolar lines* D_{E_1} and D_{E_2}. Any point I_1 of the epipolar line D_{E_1} has its potential matches on D_{E_2} and vice versa.

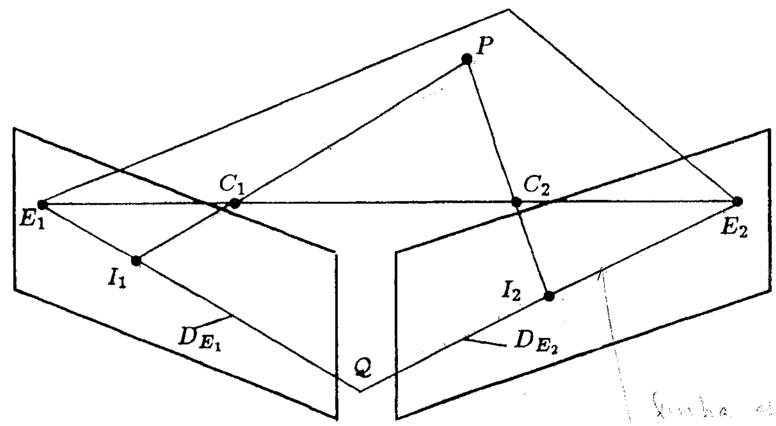

Figure 3.2
The geometry of binocular stereo vision.

We see as well that any epipolar line of image 2 is the image of a line passing through C_1. From this it follows that the epipolar lines of image 2 form a ***bundle of lines*** with center E_2, the image of C_1 in camera 2 (cf. Fig. 3.3). E_2 is called the *epipole* of image 2. The situation is symmetric, and E_1, the image of C_2 in camera 1 is the epipole of image 1. We may also consider E_1 and E_2 as the intersections of the line C_1C_2 with the image planes $\mathcal{P}_1$ and $\mathcal{P}_2$ respectively.

3.4.2 Computation of epipoles

Once the perspective transformation matrices T_1 and T_2 of cameras 1 and 2 are known, the computation of the epipoles is as follows. We must first determine the coordinates $(x_{C_i}, y_{C_i}, z_{C_i})$ of the optical centers C_i. These are obtained by solving the system

$$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = T_i \begin{pmatrix} x_{C_i} \\ y_{C_i} \\ z_{C_i} \\ 1 \end{pmatrix}$$

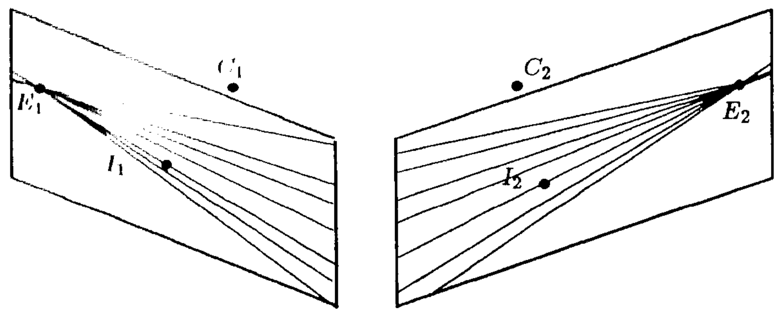

Figure 3.3
The bundle of epipolar lines: general case.

The projective coordinates of the epipole E_2 are then obtained by

$$E_2^* = \begin{pmatrix} U_{E_2} \\ V_{E_2} \\ S_{E_2} \end{pmatrix} = T_2 \begin{pmatrix} x_{C_1} \\ y_{C_1} \\ z_{C_1} \\ 1 \end{pmatrix} \tag{3.4.2}$$

Epipoles with finite coordinates If S_{E_2} is non-zero, the image coordinates of E_2 are determined from the projective coordinates by

$$\begin{pmatrix} u_{E_2} \\ v_{E_2} \end{pmatrix} = \begin{pmatrix} U_{E_2}/S_{E_2} \\ V_{E_2}/S_{E_2} \end{pmatrix} \tag{3.4.3}$$

The computation of the epipole E_1 of image 1 proceeds similarly.

One epipole at infinity If S_{E_2} is zero, this means the epipole occurs at infinity. This happens when the optical center of camera 1 is in the focal plane of camera 2. In this case, the epipolar lines form a bundle of parallel lines in image 2 and the situation is as in figure 3.4.

Two epipoles at infinity The very special case shown in figure 3.5 which corresponds to bundles of parallel epipolar lines in both images simultaneously, can only happen when both S_{E_1} and S_{E_2} are zero, that is, when the line C_1C_2 is parallel to the image planes of both cameras.[1] In practice, this exceptional situation is very difficult to obtain, but, as we see below, it is possible to easily arrange for such a configuration by an *a posteriori* transformation of the images called *rectification*.

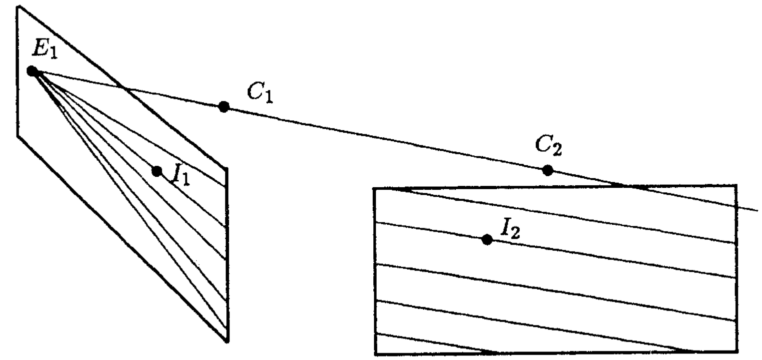

Figure 3.4
The bundle of parallel epipolar lines in one of the images.

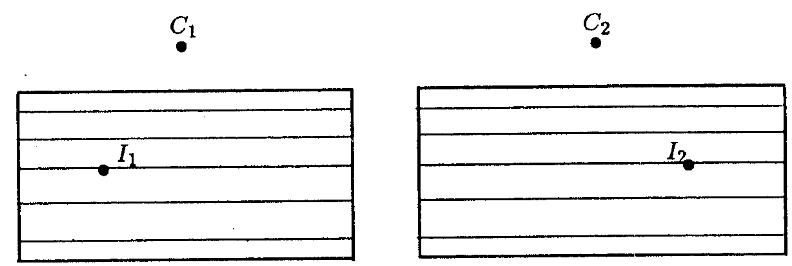

Figure 3.5
Bundles of parallel epipolar lines in both images.

[1] In figure 3.5, we have also made the axis u parallel to the direction C_1C_2 in both images.

3.4.3 Parametric equation of epipolar lines

Here, we analytically compute the equation of the epipolar line D_{E_2} of image 2 associated with the image point I_1 of image 1 whose coordinates are (u_1, v_1). Let us again consider the system 3.3.1 relating point I_1 to scene points P of which it may potentially be the image, and rewrite it in the form:

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

where the index j of t_i^j refers to camera j.

These are the equations of the two planes whose intersection is the line $C_1 I_1$. The direction vector n of this line is the vector product of the normal vectors of the two planes:

$$n = (t_1^1 - u_1\, t_3^1) \times (t_2^1 - v_1\, t_3^1)$$

i.e.,

$$n = u_1\, t_2^1 \times t_3^1 + v_1\, t_3^1 \times t_1^1 + t_1^1 \times t_2^1 \tag{3.4.4}$$

The parametric equation of the physical line $C_1 I_1$ is $P = C_1 + \lambda n$. Thus, its image in camera 2 is formed by the set of points I_2 whose projective coordinates I_2^* satisfy:

$$I_2^* = T_2 \begin{pmatrix} C_1 + \lambda\, n \\ 1 \end{pmatrix}$$

Letting

$$F_2^* = T_2'\, n \tag{3.4.5}$$

where T_2' is the 3x3 submatrix extracted from T_2 by dropping the last column, we have the parametric equation of the epipolar line *in projective coordinates:*

$$I_2^* = E_2^* + \lambda\, F_2^*$$

The parametric equation in image coordinates is thus given by

$$u_2 = \frac{U_{E_2} + \lambda\, U_{F_2}}{S_{E_2} + \lambda\, S_{F_2}} \tag{3.4.6}$$

$$v_2 = \frac{V_{E_2} + \lambda\, V_{F_2}}{S_{E_2} + \lambda\, S_{F_2}} \tag{3.4.7}$$

3.4.4 Computation in practice

Which parameters are necessary to compute the epipolar line associated with an arbitrary image point I_1? We distinguish two cases:

The epipole E_2 is defined ($S_{E_2} \neq 0$) In this case, the epipolar line D_{E_2} passes through the epipole E_2 with coordinates defined in the image by the systems of equations 3.4.2 and 3.4.3. The coordinates of the epipole do not depend on I_1. We may compute them once and for all at calibration from the matrices T_1 and T_2 alone, and thus store two image coordinates.

It only remains to compute the direction vector D_{E_2}. This may be determined, for example, by computing derivatives of equations 3.4.6 and 3.4.7 with respect to λ. We thus get the direction vector:

$$\begin{pmatrix} \Delta u_2 \\ \Delta v_2 \end{pmatrix} = \begin{pmatrix} U_{F_2} S_{E_2} - U_{E_2} S_{F_2} \\ V_{F_2} S_{E_2} - V_{E_2} S_{F_2} \end{pmatrix} \tag{3.4.8}$$

Looking at equations 3.4.4 and 3.4.5, we see that the direction vector is computed as an affine function of the coordinates (u_1, v_1) of I_1. It is sufficient to set

$$M_{21} = \begin{pmatrix} S_{E_2} & 0 & -U_{E_2} \\ 0 & S_{E_2} & -V_{E_2} \end{pmatrix} T_2' \,[t_2^1 \times t_3^1 \quad t_3^1 \times t_1^1 \quad t_1^1 \times t_2^1]$$

to obtain

$$\begin{pmatrix} \Delta u_2 \\ \Delta v_2 \end{pmatrix} = M_{21} \begin{pmatrix} u_1 \\ v_1 \\ 1 \end{pmatrix}$$

The matrix M_{21} is a 2×3 matrix *independent* of I_1. It is thus sufficient to compute it from only T_1 and T_2 once at calibration and to store its 6 parameters.

> We mention the special case of $E_2^* \times F_2^* = 0$. Note that the direction vector of the epipolar — given by equation 3.4.8 — is zero. In this case, the points I_1, C_1 and C_2 are aligned, which means that I_1 is placed on E_1. As a consequence, the epipolar line D_{E_2} is reduced to the point E_2, the image of E_1.

Similarly, we compute and store the coordinates of the epipole E_1 and a matrix M_{12} for image 1.

The epipole E_2 is undefined ($S_{E_2} = 0$) In this case, the epipole E_2 is considered to be at infinity. Equations 3.4.8 show that the direction vector no longer depends on I_1. As a result, all epipolar lines are parallel to the direction vector

$$\begin{pmatrix} \Delta u_2 \\ \Delta v_2 \end{pmatrix} = \begin{pmatrix} U_{E_2} \\ V_{E_2} \end{pmatrix}$$

which may be computed once and for all at calibration (two parameters).

In addition, the line D_{E_2} passes through the vanishing point

$$\left(\frac{U_{F_2}}{S_{F_2}}, \frac{V_{F_2}}{S_{F_2}}\right)$$

(see next section). The projective coordinates of this point are computed by an affine transformation of the coordinates (u_1, v_1) of I_1. We set

$$N_{21} = T_2' \, [t_2^1 \times t_3^1 \quad t_3^1 \times t_1^1 \quad t_1^1 \times t_2^1]$$

in order to obtain

$$F_2^* = N_{21} \begin{pmatrix} u_1 \\ v_1 \\ 1 \end{pmatrix}$$

N_{21} is a 3×3 matrix *independent* of I_1. It is sufficient to compute it only once at calibration from T_1 and T_2 alone and then to store its 9 parameters.

> Note that $S_{F_2} = 0$ is a degenerate case, i.e., when the point I_1 is in the focal plane of camera 2. As C_1 is already in this focal plane (since $S_{E_2} = 0$), the epipolar plane Q is parallel to the image plane $\mathcal{P}_2$ and the epipolar line D_{E_2} is undefined.

Similarly, we compute and store the coordinates of the epipole E_1 and store a matrix N_{12} for image 1.

3.5 Disparity

3.5.1 Definition

To determine the matching point for I_1 in the image plane $\mathcal{P}_2$, we can now compute the epipolar line D_2 containing I_2. The position of I_2 in image 2 can thus be measured relative to the position of I_1 using a single parameter, for example, the difference in abscissas of the two image points[2] :

Definition 3.1 (disparity) *The disparity between a pair of homologous points (I_1, I_2) is defined by the difference*

$$\delta = u_2 - u_1$$

Finding I_2 now corresponds to the determination of the disparity δ with I_1. In practice, the possible values of δ are limited

1. because the observed points are necessarily in the half-space situated in front of the image planes and the optical centers of each camera;
2. by the dimensions of the images;
3. by the dimensions of the observed scene, when finite.

In the following sections, we will make precise how to impose these constraints, first for a special case and then for the general case.

3.5.2 The interval of allowable disparities

Special case Suppose that the cameras are identical, that their image planes are identical, and that the axes of abscissas in the two images are identical and parallel to C_1C_2. Consider figure 3.6, seen from the epipolar plane defined by I_1, C_1 and C_2. Let the distance between the optical axes be b and the distance between the optical centers and the epipolar line I_1I_2 be f,[3] and, in addition, let u_1 and u_2 be measured relative to the projections of C_1 and C_2 respectively. Then we can compute the distance ρ between P and the focal plane C_1C_2 as

$$\rho = \frac{b\,f}{u_2 - u_1} = \frac{b\,f}{\delta}$$

[2] Providing that the epipolar line D_2 is not vertical; more generally, we must measure the disparity along the epipolar lines, for example by the difference $\delta = E_2I_2 - E_1I_1$.

[3] f is the focal distance if the epipolar plane is orthogonal to the image planes.

In this special case, the disparity characterizes the distance to the cameras very simply, and we have the additional properties

$$\lim_{\rho\to\infty} \delta = 0\ ; \quad \lim_{\delta\to\infty} \rho = 0 \qquad (3.5.9)$$

We need only define an interval $\rho \in [\rho_{min}, \rho_{max}]$ of possible distances from point P to the two cameras in order to compute the interval of allowable disparities $\delta \in [\delta_{min}, \delta_{max}]$ independent of the position of the point I_1.

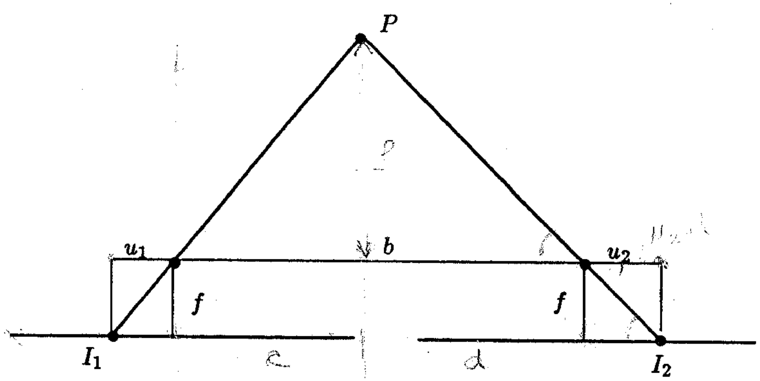

Figure 3.6
Disparity and distance: special case.

General case Suppose now that the geometry of the cameras is arbitrary and consider figures 3.7 and 3.8 seen from the epipolar plane defined by I_1, C_1 and C_2.

Given point I_1, we try to relate the variation of the position of I_2 to the distance between the physical point P and the optic center of camera 1. To do so, we consider the parametric equation of the line C_1I_1 in the form

$$C_1P = \frac{\rho}{\alpha} n$$

where n is the direction vector of the line computed by equation 3.4.4 and

$$\alpha = \|n\|$$

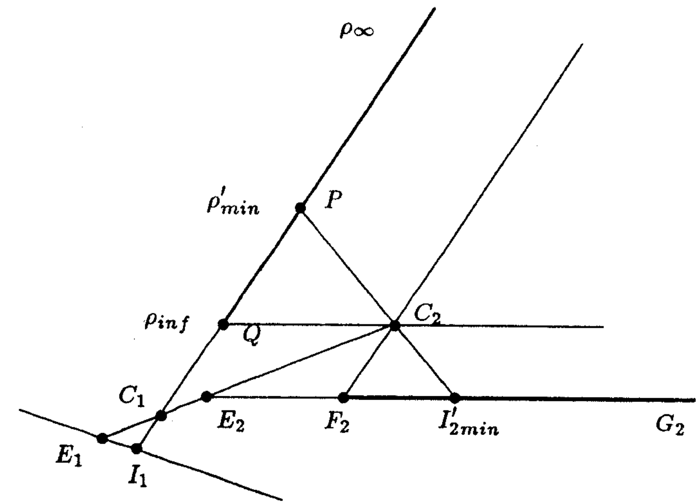

Figure 3.7
The interval of allowable disparities: first case.

is its Euclidean norm. Then ρ is the Euclidean distance between P and the optical center of camera 1.

We next compute the image I_2 of P as in section 3.4.3, which leads to equations quite similar to 3.4.6 and 3.4.7:

$$u_2 = \frac{\alpha\, U_{E_2} + \rho\, U_{F_2}}{\alpha\, S_{E_2} + \rho\, S_{F_2}} \qquad (3.5.10)$$

$$v_2 = \frac{\alpha\, V_{E_2} + \rho\, V_{F_2}}{\alpha\, S_{E_2} + \rho\, S_{F_2}} \qquad (3.5.11)$$

In these equations, F_2 is the image point given by equation 3.4.5, and E_2 is the epipole of image 2 given by equation 3.4.2.

We make the following observations:

- If we orient the vector n' of I_1 toward C_1, the half-space observable

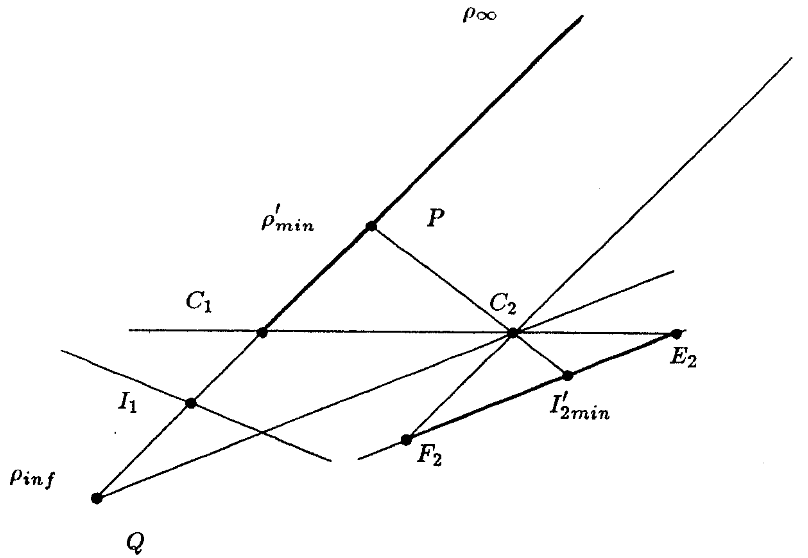

Figure 3.8
The interval of allowable disparities: second case.

by camera 1 can be determined by setting $\rho > 0$.

- When the distance ρ approaches infinity, the point I_2 approaches a limiting point F_2 with coordinates

$$u_{F_2} = \frac{U_{F_2}}{S_{F_2}}$$
$$v_{F_2} = \frac{V_{F_2}}{S_{F_2}}$$

This point F_2 is the vanishing point associated with the direction n' of the line C_1I_1. The image of any line segment in the scene whose direction is n' contains the point F_2. (cf. Fig. 3.7 and Fig. 3.8).

- When ρ takes the value $\rho_{inf} = -\alpha\, S_{E_2}/S_{F_2}$, the coordinates of I_2

become infinite. This corresponds to the movement of P through the focal plane of camera 2.

  - If $\rho_{inf} > 0$, the optical center of camera 1 lies between the image plane and the focal plane of camera 2. Then we have the situation of figure 3.7, and it becomes necessary to impose a minimum distance $\rho_{min} > \rho_{inf}$ to include only those points P which are observable by the two cameras. We call G_2 the limit point when ρ approaches ρ_{inf} from above.
  - When $\rho_{inf} \leq 0$, we have the situation of figure 3.8 and the coordinates of the image point I_2 remain finite when ρ decreases from $+\infty$ to 0.

The preceding remarks can be summarized as follows:

- The interval of possible variation of the distance ρ is of the form $[\rho_{min}, +\infty[$ with $\rho_{min} \geq \max(0, \rho_{inf})$.
- The interval of variation of I_2 is defined by E_2F_2 or G_2F_2 according to whether $\rho_{min} = 0$ or $\rho_{min} = \rho_{inf}$ respectively.

These intervals are represented as thick lines in figures 3.7 and 3.8.

In practice, objects are observable only when they are at a strictly positive distance from the optical center of the camera. We thus generally choose a distance ρ'_{min} included in the above theoretic interval $[\rho_{min}, +\infty[$, which defines a smaller interval of possible positions F_2, I'_{2min} for I_2. This was shown in the two preceding figures.

Similarly, a constrained environment such as an office interior generally imposes an upper limit to the maximum distance between an object and the camera. We can thus further reduce the interval of possible positions of I_2.

3.5.3 Disparity gradient

Another important constraint on stereo matching is the disparity gradient.

Definition 3.2 (disparity gradient) *the disparity gradient γ is defined by the partial derivative of the function $\delta = f(\rho, u_1, v_1)$ with respect* *to ρ. It is computed at a point (ρ, u_1, v_1) as*

$$\gamma(\rho, u_1, v_1) = \frac{\partial\delta}{\partial\rho}(\rho, u_1, v_1)$$

A similar definition applies to a point I_2 of image 2.

3.5.4 Computation in practice

We only compute the interval of allowable disparities and the disparity gradient for the center of a set of square windows forming a regular partition of the images. For the matching algorithms we only retain, for each of the regions of the image, the interval of allowable disparities and a table of the disparity gradient over this interval.

For the center of three of the 16 regions of image 1, table 3.1 shows

- the values of allowable disparity (the minimum and maximum distances ρ were chosen equal to 180 cm and 1995 cm respectively);
- the disparity gradient (multiplied by 20);
- the corresponding distance ρ.

  The disparity gradient is tabulated so that, between two successive values of δ, the variation in disparity gradient (multiplied by 20) is less than a given threshold $\Delta\gamma_{max}$. In the example, we have chosen $\Delta\gamma_{max} = 1.5$ pixel.

The first six lines of the table concern point I_1 with coordinates (64,64). This is the center of the first region of image 1. We see that the interval of allowable disparities of a homologous point I_2 is approximately [-19 , 38] pixels, which means that the abscissa of I_2 can vary over the interval [64 - 19 , 64 + 38] pixels.

Note that this corresponds to a distance of between 185 and 485 cm from the observed point, and that this interval is smaller than the originally specified interval ([180, 1995] cm). This is due to the limited size of image 2. In fact, outside the interval of distances [185, 485] cm, the ordinate of image point I_2 is no longer in the interval of [0, 512] pixels.

Observe as well that the gradient disparity (×20) is 9.41 pixels/cm at 180 cm, which means that at that distance (to within a linear approximation) a depth variation of 20 cm along the segment C_1I_1 becomes a disparity variation of 9.41 pixels in image 2.

Region center (pixels)	*Disparity δ* (pixels)	*Gradient $20\times\gamma$* (pixels/cm)	*Distance ρ* (cm)
(64, 64)	-18.65	1.46	485
	-2.68	3.01	335
	9.58	4.57	270
	20.47	6.22	230
	29.34	7.75	205
	38.10	9.41	185
(320, 64)	-27.85	1.24	555
	-10.54	2.75	370
	2.48	4.26	295
	13.94	5.86	250
	24.09	7.49	220
	34.85	9.42	195
	42.67	10.97	180
(448,192)	-50.00	0.10	1995
	-19.11	1.62	495
	-2.41	3.19	350
	10.44	4.75	285
	21.63	6.35	245
	32.66	8.16	215
	41.83	9.82	195
	50.00	11.43	180

Table 3.1
Table of allowable disparities and disparity gradients for three of the 16 regions of image 1.

Looking at the data relative to each of the three regions of the table, one sees a large variation in the intervals of allowable disparity. One also sees that the disparity gradient varies by a factor of 100 with distance.

These observations confirm the importance of locally computing the intervals of allowable disparity and disparity gradient in order to tune decision thresholds used at subsequent stages of the stereo vision algorithms.