17

Passive Navigation & Structure from Motion

In this chapter we investigate the problem of passive navigation using optical flow information. We wish to determine the motion of a camera from time-varying imagery or, in the discrete case, from an image sequence. Our assumption is that the camera is moving with respect to a fixed environment. If there is more than one object with independent motion, then we shall assume that the image has been segmented and that we can concentrate on one region corresponding to a single object. Curiously, if there is a translational component to the motion, then we obtain the shape of the surface as a by-product of the recovery of the motion parameters.

We discuss three approaches that have been taken to this problem. We then focus on a least-squares approach that takes into account all the information in the image. Closed-form solutions are derived for the cases of purely rotational and purely translational motion, and the set of nonlinear equations that must be solved iteratively in the general case is displayed. In the exercises a method is outlined that solves the problem without computing the optical flow as an intermediate product.

Recovering motion and surface shape from time-varying imagery is useful in guiding a moving platform through a known environment. The shapes recovered can be used for recognition by one of the techniques considered in earlier chapters.

17.1 Recovering the Motion of the Observer

Suppose we are viewing a film. We wish to determine the motion of the camera from the sequence of images, assuming that the instantaneous velocity of the brightness patterns, also called the *optical flow*, is known at each point in the image. Several schemes for recovering the observer's motion have been suggested. These approaches can be classified into three categories:

- The discrete approach.
- The differential approach.
- The least-squares approach.

In the discrete approach, information about the movement of brightness patterns at only a few points is used tc determine the motion of the camera. In particular, using such an approach, we attempt to identify and match discrete points in a sequence of images. Of interest in this case is the photogrammetric problem of determining the minimum number of points from which the motion can be calculated for a given number of images. This approach requires that we track features, or identify corresponding features in images taken at different times. It can be shown that, in general, seven points are sufficient to determine the motion uniquely. These points must satisfy a constraint on their position, but the constraint is a weak one. It is even possible, if more points are available, to write a set of linear equations in the unknown parameters of the motion.

In the differential approach, the first and second spatial partial derivatives of the optical flow are used to compute the motion of a camera. It has been claimed that it is sufficient to know the optical flow and both its first and second derivatives at a single point to determine the motion uniquely. This turns out to be incorrect, except for a special case. Furthermore, noise in the measured optical flow is accentuated by differentiation.

In the least-squares approach, the whole optical flow field is used. A major shortcoming of both the discrete and differential approaches is that neither allows for errors in the optical flow data. This is why we choose the least-squares approach to devise a technique to determine the motion of the camera from the measured optical flow. The algorithm takes the abundance of available data into account and is robust enough to allow numerical implementation.

17.2 Technical Prerequisites

In this section we review the equations describing the relation between the motion of a camera and the optical flow that motion generates. We can assume either a fixed camera in a changing environment or a moving camera in a static environment. Let us assume a moving camera in a static environment. A coordinate system is fixed with respect to the camera, with the Z-axis pointing along the optical axis. Any rigid body motion can be resolved into two components, a translation and a rotation about an axis through the origin. We shall denote the translational component of the motion of the camera by $\mathbf{t}$ and its angular velocity by $\boldsymbol{\omega}$. Let the instantaneous coordinates of a point P in the environment be $(X,Y,Z)^T$. (Note that here $Z>0$ for points in front of the imaging system.)

Let $\mathbf{r}$ be the column vector $(X,Y,Z)^T$, where ^{T} denotes the transpose. Then the velocity of P with respect to the XYZ coordinate system is

$$\mathbf{V} = -\mathbf{t} - \boldsymbol{\omega} \times \mathbf{r}.$$

If we define the components of $\mathbf{t}$ and $\boldsymbol{\omega}$ as

$$\mathbf{t} = (U,V,W)^T \quad \text{and} \quad \boldsymbol{\omega} = (A,B,C)^T,$$

we can rewrite this equation in component form as

$$\begin{aligned}\dot{X} &= -U - BZ + CY,\\ \dot{Y} &= -V - CX + AZ,\\ \dot{Z} &= -W - AY + BX,\end{aligned}$$

where the dot denotes differentiation with respect to time.

The *optical flow* at each point in the image plane is the instantaneous velocity of the brightness pattern at that point. Let (x,y) denote the coordinates of a point in the image plane. Here we assume perspective projection between an object point P and the corresponding image point p; thus the coordinates of p are

$$x = \frac{X}{Z} \quad \text{and} \quad y = \frac{Y}{Z}.$$

The optical flow at a point (x,y), denoted by (u,v), is

$$u = \dot{x} \quad \text{and} \quad v = \dot{y}.$$

Differentiating the equations for x and y with respect to time and using the derivatives of X, Y, and Z, we obtain the following equations for the optical flow:

$$\begin{aligned}u &= \frac{\dot{X}}{Z} - \frac{X\dot{Z}}{Z^2} = \left(-\frac{U}{Z} - B + Cy\right) - x\left(-\frac{W}{Z} - Ay + Bx\right),\\ v &= \frac{\dot{Y}}{Z} - \frac{Y\dot{Z}}{Z^2} = \left(-\frac{V}{Z} - Cx + A\right) - y\left(-\frac{W}{Z} - Ay + Bx\right).\end{aligned}$$

We can write these equations in the form

$$u = u_t + u_r \quad \text{and} \quad v = v_t + v_r,$$

where (u_t, v_t) denotes the translational component of the optical flow and (u_r, v_r) the rotational component:

$$\begin{aligned}u_t &= \frac{-U + xW}{Z} \quad \text{and} \quad u_r = A\,xy - B\,(x^2+1) + C\,y,\\ v_t &= \frac{-V + yW}{Z} \quad \text{and} \quad v_r = A\,(y^2+1) - B\,xy - C\,x.\end{aligned}$$

So far we have considered a single point P. To define the optical flow globally we assume that P lies on a surface defined by a function $Z(X,Y)$ that is positive for all values of X and Y. With any surface and any motion of a camera we can therefore associate a certain optical flow, and we say that the surface and the motion generate this optical flow.

Optical flow, therefore, depends upon the six parameters of motion of the camera and upon the surface whose images are analyzed. Can all these unknowns be uniquely recaptured solely from optical flow? Strictly speaking, the answer is no. To see this, consider a surface S_2 that is a dilation by a factor k of a surface S_1. Furthermore, let two motions M_1 and M_2 have the same rotational component, and let their translational components be proportional to each other with the same proportionality factor k (we say that M_1 and M_2 are *similar*). Then the optical flow generated by S_1 and M_1 is the same as the optical flow generated by S_2 and M_2. This follows directly from the definition of optical flow given above.

Determining the motion of a camera from the optical flow is much easier if the motion is purely translational or purely rotational. The next two sections deal with these two special cases. We then turn to the case in which no a priori assumptions about the motion of the camera are made.

17.3 The Translational Case

In this section we discuss the case in which the motion of the camera is purely translational. As before, let $\mathbf{t} = (U,V,W)^T$ be the velocity of the camera. Then the following equations hold:

$$u_t = \frac{-U + xW}{Z} \quad \text{and} \quad v_t = \frac{-V + yW}{Z}.$$

17.3.1 Similar Surfaces and Similar Motions

We want to show that if two purely translational motions generate the same optical flow, the two surfaces are similar and the two camera motions are similar. Let Z_1 and Z_2 be two surfaces and let $\mathbf{t}_1 = (U_1, V_1, W_1)^T$ and $\mathbf{t}_2 = (U_2, V_2, W_2)^T$ define two different motions of a camera such that Z_1 and $\mathbf{t}_1$ and Z_2 and $\mathbf{t}_2$ generate the same optical flow, that is,

$$u = \frac{-U_1 + xW_1}{Z_1} \quad \text{and} \quad v = \frac{-V_1 + yW_1}{Z_1},$$
$$u = \frac{-U_2 + xW_2}{Z_2} \quad \text{and} \quad v = \frac{-V_2 + yW_2}{Z_2}.$$

Eliminating Z_1, Z_2, u, and v from these equations, we obtain

$$\frac{-U_1 + xW_1}{-V_1 + yW_1} = \frac{-U_2 + xW_2}{-V_2 + yW_2}.$$

We can rewrite this as

$$(-U_1 + xW_1)(-V_2 + yW_2) = (-U_2 + xW_2)(-V_1 + yW_1),$$

or

$$U_1V_2 - xV_2W_1 - yU_1W_2 + xyW_1W_2 = U_2V_1 - xV_1W_2 - yU_2W_1 + xyW_2W_1.$$

Since we are assuming that Z_1 and $\mathbf{t}_1$ and Z_2 and $\mathbf{t}_2$ generate the same optical flow, the above equation must hold for all x and y. Therefore the following must hold:

$$U_1V_2 = U_2V_1,$$
$$-V_2W_1 = -V_1W_2,$$
$$-U_1W_2 = -U_2W_1.$$

These equations can be rewritten in ratio form as

$$U_1 : V_1 : W_1 = U_2 : V_2 : W_2,$$

from which it follows that Z_2 is a dilation of Z_1. It is clear that the scaling factor between Z_1 and Z_2 (or equivalently between $\mathbf{t}_1$ and $\mathbf{t}_2$) cannot be recovered from the optical flow, regardless of the number of points at which the flow is known. We shall say that the motion of the camera is uniquely determined if it has been determined up to a constant scaling factor.

17.3.2 A Least-Squares Formulation

In general, the directions of the optical flow at two points in the image plane uniquely determine the purely translational motion of a camera. There is, however, a drawback to utilizing so little of the available information. The optical flow we measure is corrupted by noise, and we would like to develop a robust method that takes this into account. Thus we suggest using a least-squares method to determine the movement parameters and the surface (that is, the best fit with respect to some norm).

For the following we assume that the image plane is the rectangle $x \in [-w, w]$ and $y \in [-h, h]$. The same method applies if the image has some other shape. (In fact, it can be used on subimages corresponding to individual objects, in the situation where the environment contains several objects that can move with respect to one another.) Usually, there is a lower limit to the distance between the objects and the camera, so we can assume that $1/Z$ is a bounded function. Moreover, most scenes will consist of a number of cohesive objects with continuous surfaces, so that depth will be continuous "almost everywhere." Specifically, let us assume that the set of points where $1/Z$ is discontinuous is of *measure zero*. (This means that the integral of a characteristic function, which equals one in these places and zero elsewhere, is zero.) This condition ensures that all necessary integrations can be carried out.

We wish to minimize the expression

$$\iint \left(\left(u - \frac{-U + xW}{Z}\right)^2 + \left(v - \frac{-V + yW}{Z}\right)^2 \right) dx\, dy.$$

In this case, then, we determine the best fit with respect to the ML_2 norm, which is defined as

$$\|f(x, y)\| = \iint (f(x, y))^2 \, dx\, dy.$$

The steps in the least-squares method are as follows: First, we determine the value of Z that minimizes the integrand at every point (x, y); then we determine the values of U, V, and W that minimize the integral.

It is convenient to define

$$\alpha = -U + xW \quad \text{and} \quad \beta = -V + yW.$$

Note that the expected flow, given U, V, and W, is simply

$$\overline{u} = \frac{\alpha}{Z} \quad \text{and} \quad \overline{v} = \frac{\beta}{Z}.$$

Then we can rewrite the integral above as

$$\iint \left(\left(u-\frac{\alpha}{Z}\right)^2+\left(v-\frac{\beta}{Z}\right)^2\right)\,dx\,dy.$$

We proceed now with the first step of our minimization method. Differentiating the integrand with respect to Z and setting the resulting expression equal to zero yields

$$\left(u-\frac{\alpha}{Z}\right)\frac{\alpha}{Z^2}+\left(v-\frac{\beta}{Z}\right)\frac{\beta}{Z^2}=0.$$

Therefore we can write Z as

$$Z=\frac{\alpha^2+\beta^2}{u\alpha+v\beta}.$$

This equation, by the way, imposes a constraint on U, V, and W, since Z must be positive. We do not make use of this except to help us choose between two opposite solutions for the translational velocity later on. Note that now

$$u-\frac{\alpha}{Z}=+\beta\frac{u\beta-v\alpha}{\alpha^2+\beta^2}\qquad\text{and}\qquad v-\frac{\beta}{Z}=-\alpha\frac{u\beta-v\alpha}{\alpha^2+\beta^2},$$

and we can therefore rewrite the integral above as

$$\iint \frac{(u\beta-v\alpha)^2}{\alpha^2+\beta^2}\,dx\,dy.$$

It should be clear that uniformly scaling U, V, and W does not change the value of the integral. This is a reflection of the fact that we can determine the motion parameters only up to a constant factor.

Before proceeding with the second step, we give a geometrical interpretation of what we have done so far. Suppose that the motion parameters U, V, and W are given. At any given point (x_0, y_0), optical flow depends not only upon the motion parameters but also upon the value Z_0 of Z at that point. However, the direction of (u, v) does not depend upon Z_0. The point (u, v) must lie along the line L in the uv-plane defined by the equation $u\beta - v\alpha = 0$. Let the measured optical flow at (x_0, y_0) be denoted (u_m, v_m), and let the closest point on the line L be (u_a, v_a). This corresponds to a particular Z_a. The remaining error is the distance between the point (u_m, v_m) and the line L. The square of this distance is given by the integrand of the integral above.

For the second step, we differentiate the integral with respect to U, V, and W and set the resulting expressions equal to zero:

$$\iint \frac{\beta(u\beta-v\alpha)(u\alpha+v\beta)}{(\alpha^2+\beta^2)^2}\,dx\,dy=0,$$

$$-\iint \frac{\alpha(u\beta-v\alpha)(u\alpha+v\beta)}{(\alpha^2+\beta^2)^2}\,dx\,dy=0,$$

$$\iint \frac{(y\alpha-x\beta)(u\beta-v\alpha)(u\alpha+v\beta)}{(\alpha^2+\beta^2)^2}\,dx\,dy=0.$$

If we introduce the abbreviation

$$K=\frac{(u\beta-v\alpha)(u\alpha+v\beta)}{(\alpha^2+\beta^2)^2},$$

the equations can be rewritten as

$$\iint \big((-V+yW)K\big)\,dx\,dy=0,$$

$$-\iint \big((-U+xW)K\big)\,dx\,dy=0,$$

$$\iint \big((-yU+xV)K\big)\,dx\,dy=0.$$

The sum of U times the first integral, V times the second integral, and W times the third integral is identically zero. Thus the three equations are linearly dependent. This is to be expected, for if

$$f(kU, kV, kW) = f(U, V, W),$$

where f is a differentiable function and k a constant, then

$$U\frac{\partial f}{\partial U}+V\frac{\partial f}{\partial V}+W\frac{\partial f}{\partial W}=0.$$

The result is also consistent with the fact that only two equations are needed, since the translational velocity can be determined only up to a constant factor. Unfortunately, the equations are nonlinear in U, V, and W, and we are not able to show that they have a unique (up to a constant scaling factor) solution.

17.3.3 Using a Different Norm

There is a way, however, to devise a least-squares method that allows us to display a closed-form solution for the motion parameters. Instead of minimizing the integral above, we try to minimize the expression

$$\iint \left(\left(u - \frac{-U + xW}{Z}\right)^2 + \left(v - \frac{-V + yW}{Z}\right)^2\right) (\alpha^2 + \beta^2)\, dx\, dy,$$

obtained by multiplying the integrand by $\alpha^2 + \beta^2$. Then we apply the same least-squares method as before. When the measured optical flow is not corrupted by noise, both integrals can be made equal to zero by substituting the correct motion parameters. We thus obtain the same solution for the motion parameters. If the measured optical flow is not exact, then using the new integral for our minimization yields the best fit with respect not to the ML_2 norm, but to another norm that we shall call the $ML_{\alpha\beta}$ norm, namely

$$\|f(x, y)\|_{\alpha\beta} = \iint \big(f(x, y)\big)^2 (\alpha^2 + \beta^2)\, dx\, dy.$$

What we have here is a minimization in which the error contributions are weighted, greater importance being given to points where the optical flow velocity is larger. This is most appropriate when the measurement of larger velocities is more accurate.

Which norm gives the best results depends on the properties of the noise in the measured optical flow. The first norm is better suited to the situation in which the noise in the measurements is independent of the magnitude of the optical flow. Note also that if we really want the minimum with respect to the ML_2 norm, we can use the results of the minimization with respect to the $ML_{\alpha\beta}$ norm as starting values in a numerical process.

We now apply our least-squares method to the case in which the norm is chosen to be $ML_{\alpha\beta}$. First, we determine Z by differentiating the integrand with respect to Z and setting the result equal to zero. We again obtain

$$\left(u - \frac{\alpha}{Z}\right) \frac{\alpha}{Z^2} + \left(v - \frac{\beta}{Z}\right) \frac{\beta}{Z^2} = 0,$$

from which it follows that

$$Z = \frac{\alpha^2 + \beta^2}{u\alpha + v\beta}.$$

We therefore want to minimize

$$\iint (u\beta - v\alpha)^2\, dx\, dy.$$

If we call this integral $g(U, V, W)$, then, since

$$u\beta - v\alpha = (vU - uV) - (xv - yu)W,$$

we have

$$g(U, V, W) = aU^2 + bV^2 + cW^2 + 2dUV + 2eVW + 2fWU,$$

where

$$a = \iint v^2\, dx\, dy,$$

$$b = \iint u^2\, dx\, dy,$$

$$c = \iint (xv - yu)^2\, dx\, dy,$$

$$d = -\iint uv\, dx\, dy,$$

$$e = \iint u(xv - yu)\, dx\, dy,$$

$$f = -\iint v(xv - yu)\, dx\, dy.$$

Now $g(U, V, W)$ cannot be negative, and $g(U, V, W) = 0$ for $U = V = W = 0$. Thus a minimum can be found by inspection, but it is not what we might have hoped for! In fact, to determine the translational velocity using our least-squares method, we must solve the following homogeneous equation for $\mathbf{t}$:

$$\mathbf{G}\,\mathbf{t} = 0,$$

where

$$\mathbf{G} = \begin{pmatrix} a & d & f \\ d & b & e \\ f & e & c \end{pmatrix}.$$

This clearly has a solution other then zero if and only if the determinant of $\mathbf{G}$ is zero. Then the three equations are linearly dependent and $\mathbf{t}$ can be determined up to a constant factor. In general, however, because the data are corrupted by noise, $g(U, V, W)$ cannot be made equal to zero for nonzero translational velocity, and so $\mathbf{t} = (0, 0, 0)^T$ will be the only solution. To see this another way, note that g has the form

$$g(kU, kV, kW) = k^2 g(U, V, W),$$

where k is a constant. Clearly $g(U, V, W)$ assumes its minimum value for $U = V = W = 0$.

What we are really interested in is determining the direction of $\mathbf{t}$ that minimizes g, for a fixed length of $\mathbf{t}$. Hence we impose the constraint that $\mathbf{t}$ be a unit vector. If $\mathbf{t}$ is constrained to have unit magnitude, then the minimum value of g is the smallest eigenvalue of the matrix $\mathbf{G}$, and the value of $\mathbf{t}$ for which g assumes its minimum can be found by determining the eigenvector corresponding to this eigenvalue. This follows from the observation that g is a quadratic form that can be written as

$$g(U,V,W) = \mathbf{t}^T\mathbf{G}\mathbf{t}.$$

Note that $\mathbf{G}$ is a positive semidefinite Hermitian matrix since $a \geq 0$, $b \geq 0$, $c \geq 0$, $ab \geq d^2$, $bc \geq e^2$, and $ca \geq f^2$. (The last three inequalities follow from the Cauchy–Schwarz inequality.) Hence all the eigenvalues λ_i are real and nonnegative; they are the solutions of the third-degree polynomial

$$\begin{aligned}&\lambda^3\\ &-(a+b+c)\lambda^2\\ &+(ab+bc+ca-d^2-e^2-f^2)\lambda\\ &+(ae^2+bf^2+cd^2-abc-2def)=0.\end{aligned}$$

There is an explicit formula for the least positive root in terms of the real and imaginary parts of the roots of the quadratic resolvent of the cubic. In our case, this gives us the desired smallest root, since the roots cannot be negative. For the sake of completeness, however, various pathological cases that might come up will be discussed next, even though they are of little practical interest.

Note that $\lambda = 0$ is an eigenvalue if and only if $\mathbf{G}$ is singular, that is, if the constant term in the polynomial equals zero. In fact, if the determinant of $\mathbf{G}$ is zero, we can find a translational velocity $\mathbf{t}$ that makes g zero. It follows from a theorem in calculus that this happens only when the optical flow is correct "almost everywhere," that is, when the set of points where it is corrupted is of measure zero. The theorem states that if the integral of the square of a bounded and continuous function is zero, then the function itself is zero. Hence errors can only occur at points where the optical flow is discontinuous, and these are exactly the points where the surface defined by Z is discontinuous. (These are also the places where existing methods for computing the optical flow are subject to large errors.)

It is impossible for exactly two eigenvalues to be zero, since this would imply that the coefficient of λ in the polynomial is equal to zero, while that of λ^2 is not. That in turn would imply that $ab = d^2$, $bc = e^2$, and $ca = f^2$, while a, b, and c are not all zero. For equality to hold in the Cauchy–Schwarz inequalities, however, u and v must both be proportional to $xv - yu$. This can only be true (for all x and y in the image) if $u{=}v{=}0$. But then all six integrals become zero, and consequently all three eigenvalues are zero. This situation is of little interest, since it occurs only when the optical flow is zero everywhere. Then the velocity is also zero.

Once the smallest eigenvalue is known, it is straightforward to find the translational velocity that best matches the given data. To determine the eigenvector corresponding to an eigenvalue λ_1, we must solve the following homogeneous set of linear equations:

$$\begin{aligned}(a-\lambda_1)U + dV + fW &= 0,\\ dU + (b-\lambda_1)V + eW &= 0,\\ fU + eV + (c-\lambda_1)W &= 0.\end{aligned}$$

Because λ_1 is an eigenvalue, these equations are linearly dependent. Let us for a moment assume that all eigenvalues are distinct, that is, the rank of the matrix $(\mathbf{G} - \lambda\mathbf{I})$ is two, where $\mathbf{I}$ is the identity matrix. Then we can use any pair of equations to solve for U and V in terms of W. There are three ways to do this. To obtain a symmetric answer we add the three results:

$$\begin{aligned}U &= (b-\lambda_1)(c-\lambda_1) - f(b-\lambda_1) - d(c-\lambda_1) + e(f+d-e),\\ V &= (c-\lambda_1)(a-\lambda_1) - d(c-\lambda_1) - e(a-\lambda_1) + f(d+e-f),\\ W &= (a-\lambda_1)(b-\lambda_1) - e(a-\lambda_1) - f(b-\lambda_1) + d(e+f-d).\end{aligned}$$

Note that λ_1 will be very small if the data are good, and we might want simply to approximate the exact solution by using the above equations with λ_1 set to zero. (Then, of course, there is no need to determine the eigenvalue.) In any case, the resulting velocity can now be normalized so that its magnitude is one. There is one remaining difficulty, arising from the fact that if $\mathbf{t}$ is a solution to our minimization problem, so is $-\mathbf{t}$. Only one of these solutions will correspond to positive values of Z, however. This can be seen easily by evaluating Z at some point in the image. The case in which the two smallest eigenvalues are the same will be discussed below.

There is a simple geometrical interpretation of what we have done so far. Consider the surface defined by $g(U,V,W) = k$, where k is a constant. Note that we can always find a new coordinate system $(\widetilde{U},\widetilde{V},\widetilde{W})$ in which $g(U,V,W)$ can be written in the form

$$\lambda_1\widetilde{U}^2 + \lambda_2\widetilde{V}^2 + \lambda_3\widetilde{W}^2 = k,$$

where the λ_i ($i = 1, 2, 3$) are the three eigenvalues of the quadratic form. If the eigenvalues are all nonzero, the surface $g(U,V,W) = k$ is an ellipsoid with three orthogonal semiaxes of length $\sqrt{k/\lambda_i}$. We are particularly interested in the case where the constant k is the smallest eigenvalue. Then all three semiaxes have lengths less than or equal to one. Hence the ellipsoid lies within the unit sphere. If the two smallest eigenvalues are distinct, the unit sphere touches the ellipsoid in two places, corresponding to the largest axis. If the two smaller eigenvalues happen to be the same, however, the unit sphere touches the ellipsoid along a circle, and as a result all the velocity vectors lying in a plane spanned by two eigenvectors give equally low errors. Finally, if all three eigenvalues are equal, no direction for $\mathbf{t}$ is preferred, since the ellipsoid becomes the unit sphere.

The case in which exactly one eigenvalue is zero also has a simple geometrical interpretation. The surface defined by $g(U, V, W) = 0$ is a straight line, as can be seen easily from an examination of the equation

$$\lambda_1 \widetilde{U}^2 + \lambda_2 \widetilde{V}^2 = 0,$$

written for the case in which λ_3 is zero. (Remember that λ_1 and λ_2 are both positive.) Clearly the unit sphere intersects this line in exactly two points, one of which corresponds to positive values for Z.

The method just described can be easily implemented. To this end, the problem can be discretized. We can derive an expression similar to the above, but with the integrals approximated by sums. Our minimization method can then be applied to these sums. The resulting equations are similar to ones described in this section, with summation replacing integration. We can use the ratio of the biggest to the smallest eigenvalue, the *condition number*, as a measure of confidence in the computed velocity. The computed velocity is sensitive to errors in the measurements unless the condition number is much larger than one.

The same error integral as above is obtained when we use the ML_{Zuv} norm defined by

$$\|f(x,y)\|_{Zuv} \iint \big(f(x,y)Z(x,y)\big)^2 (u^2+v^2)\, dx\, dy.$$

Moreover, we can arrive at a similar solution by multiplying the integrand by Z^2 instead of $\alpha^2 + \beta^2$. In that case the minimization is carried out with respect to the ML_Z norm defined by

$$\|f(x,y)\|_Z = \iint \big(f(x,y)Z(x,y)\big)^2\, dx\, dy.$$

Here optical flow velocities for points that are farther away are weighted more heavily. This is most appropriate when the measurement of larger velocities is less accurate. We end up with a quadratic form similar to g, but the integrals for the six constants corresponding to a, b, c, d, e, and f are a bit more complicated. Curiously, they depend only on the direction of the optical flow at each point, not on its magnitude.

Other constraints could also be used. If we insist on $U^2 + V^2 = 1$, for example, we obtain a quadratic instead of a cubic equation, and if we use $W = 1$, only a linear equation needs to be solved. The disadvantage of these approaches is that the result is sensitive to the orientation of the coordinate axes. Clearly, in the case of exact data, we can obtain the correct solution using any of the three constraints mentioned above.

17.4 The Rotational Case

Suppose now that the motion of the camera is purely rotational. In order to determine the motion from optical flow, we again use a least-squares algorithm with the ML_2 norm described in the previous section. Recall that in this case the optical flow is

$$u_r = Axy - B(x^2+1) + Cy,$$
$$v_r = A(y^2+1) - Bxy - Cx.$$

We shall show now, in a fashion analogous to the earlier approach, that two different rotations, $\boldsymbol{\omega}_1 = (A_1, B_1, C_1)^T$ and $\boldsymbol{\omega}_2 = (A_2, B_2, C_2)^T$, cannot generate the same optical flow. If we assume the converse, the following equations must hold for all values of x and y:

$$A_1xy - B_1(x^2+1) + C_1y = A_2xy - B_2(x^2+1) + C_2y,$$
$$A_1(y^2+1) - B_1xy - C_1x = A_2(y^2+1) - B_2xy - C_2x,$$

from which we can immediately deduce that $\boldsymbol{\omega}_1 = \boldsymbol{\omega}_2$.

In general, the direction of the optical flow at two points and its magnitude at one point uniquely determine the purely rotational motion of a camera. We choose instead to minimize the following expression:

$$\iint \big((u-u_r)^2 + (v-v_r)^2\big)\, dx\, dy.$$

Because the motion is purely rotational, the optical flow does not depend upon the distance to the surface, and we can therefore omit the first step used in our method for the translational case. Thus we immediately differentiate the integral with respect to A, B, and C and set the resulting expressions to zero:

$$\iint \big((u-u_r)xy + (v-v_r)(y^2+1)\big)\, dx\, dy = 0,$$
$$\iint \big((u-u_r)(x^2+1) + (v-v_r)xy\big)\, dx\, dy = 0,$$
$$\iint \big((u-u_r)y - (v-v_r)x\big)\, dx\, dy = 0.$$

We can rewrite these equations as

$$\iint \left(uxy + v(y^2+1)\right) dx\, dy = \iint \left(u_r xy + v_r(y^2+1)\right) dx\, dy,$$
$$\iint \left(u(x^2+1) + vxy\right) dx\, dy = \iint \left(u_r(x^2+1) + v_r xy\right) dx\, dy,$$
$$\iint (uy - vx)\, dx\, dy = \iint (u_r y - v_r x)\, dx\, dy,$$

and expand them to yield

$$\overline{a}A + \overline{d}B + \overline{f}C = \overline{k},$$
$$\overline{d}A + \overline{b}B + \overline{e}C = \overline{l},$$
$$\overline{f}A + \overline{e}B + \overline{c}C = \overline{m},$$

where

$$\overline{a} = \iint \left(x^2y^2 + (y^2+1)^2\right) dx\, dy,$$
$$\overline{b} = \iint \left((x^2+1)^2 + x^2y^2\right) dx\, dy,$$
$$\overline{c} = \iint (x^2 + y^2)\, dx\, dy,$$
$$\overline{d} = -\iint \left(xy(x^2+y^2+2)\right) dx\, dy,$$
$$\overline{e} = -\iint y\, dx\, dy,$$
$$\overline{f} = -\iint x\, dx\, dy,$$

and

$$\overline{k} = \iint \left(uxy + v(y^2+1)\right) dx\, dy,$$
$$\overline{l} = -\iint \left(u(x^2+1) + vxy\right) dx\, dy,$$
$$\overline{m} = \iint (uy - vx)\, dx\, dy.$$

If we call the coefficient matrix $\mathbf{M}$ and the column vector on the right-hand side $\mathbf{n}$, we have

$$\mathbf{M}\,\boldsymbol{\omega} = \mathbf{n}.$$

Thus, provided that the matrix $\mathbf{M}$ is nonsingular, we can compute the rotation as

$$\boldsymbol{\omega} = \mathbf{M}^{-1}\mathbf{n}.$$

We show in exercise 17-9 that the matrix $\mathbf{M}$ is nonsingular in the special case of a rectangular image plane. As the extent of the image plane decreases, however, the matrix $\mathbf{M}$ becomes ill-conditioned. That is, inaccuracies in the integrals $\overline{k}$, $\overline{l}$, and $\overline{m}$ computed from the observed flow are greatly magnified. This makes sense, since we cannot expect to determine accurately the component of rotation about the optical axis when observations are confined to a small cone of directions about the optical axis.

As before, in numerical implementations of the algorithm the integrals can be approximated by sums.

17.5 General Motion

We would like now to apply a least-squares algorithm to determine the motion of a camera from optical flow data with no a priori assumptions about the motion. It is plain that a least-squares method is easiest to use when the resulting equations are linear in all the motion parameters. Unfortunately, there exists no norm that will allow us to achieve this goal. There is a norm, however, that results in equations that are linear in some of the unknowns and quadratic in the others. We again attack the minimization problem using the $ML_{\alpha\beta}$ norm under the constraint that $U^2+V^2+W^2=1$. The ensuing equations are polynomials in the unknowns U, V, W, A, B, and C and can be solved by a standard iterative procedure such as Newton's method or Bairstow's method or by an interpolation scheme such as *regula falsi*. The expression we wish to minimize is

$$\iint \left(\left(u - \left(\frac{\alpha}{Z} + u_r\right)\right)^2 + \left(v - \left(\frac{\beta}{Z} + v_r\right)\right)^2 \right) (\alpha^2 + \beta^2)\, dx\, dy.$$

The first step is to differentiate the integrand with respect to Z and set the resulting expression equal to zero, yielding

$$Z = \frac{\alpha^2 + \beta^2}{(u - u_r)\alpha + (v - v_r)\beta}.$$

We introduce the Lagrange multiplier λ and attempt to minimize

$$\iint \left((u - u_r)\beta - (v - v_r)\alpha\right)^2 dx\, dy + \lambda(U^2 + V^2 + W^2 - 1).$$

The equations we have to solve to determine the motion parameters are obtained by differentiation:

$$\iint \big((u-u_r)\beta - (v-v_r)\alpha\big)\big(-xy\beta + (y^2+1)\alpha\big)\,dx\,dy = 0,$$

$$\iint \big((u-u_r)\beta - (v-v_r)\alpha\big)\big((x^2+1)\beta - xy\alpha\big)\,dx\,dy = 0,$$

$$\iint \big((u-u_r)\beta - (v-v_r)\alpha\big)\big(y\beta + x\alpha\big)\,dx\,dy = 0,$$

$$\iint \big((u-u_r)\beta - (v-v_r)\alpha\big)(v-v_r)\,dx\,dy + \lambda U = 0,$$

$$\iint \big((u-u_r)\beta - (v-v_r)\alpha\big)(u-u_r)\,dx\,dy - \lambda V = 0,$$

$$\iint \big((u-u_r)\beta - (v-v_r)\alpha\big)\big((u-u_r)y + (v-v_r)x\big)\,dx\,dy + \lambda W = 0,$$

$$U^2 + V^2 + W^2 = 1.$$

Note that the first three of these equations are linear in A, B, and C, so that these parameters can be determined uniquely in terms of U, V, and W. Then we can determine U, V, and W from the last four equations by a numerical method. This immediately suggests an iterative scheme. To this end, we can discretize the problem and derive analogous equations in which summation of the appropriate expressions replaces integration.

In summary, our objective was to devise a method for determining the motion of a camera from optical flow, allowing for noise in the measured data. The least-squares method proposed in this chapter meets our goal and is also suitable for numerical implementation. One interesting extension explored in exercise 17-11 is the possibility of determining the motion directly from image brightness gradients without computing the optical flow.

17.6 References

The books by Hildreth, *The Measurement of Visual Motion* [1983], and by Ullman, *The Interpretation of Visual Motion* [1979], discuss the determination and interpretation of visual motion. Huang edited a collection of work presented at a NATO conference in *Image Sequence Processing and Dynamic Scene Analysis* [1983].

There has been rapid progress in this field in the last five years. The rigid-body assumption provides a powerful constraint to simplify the analysis. Most of the papers cited use it, although some allow for more than

one rigid body in the scene being viewed. The special cases of pure rotation and pure translation are particularly easy to deal with, as we discussed; see, for example, Bruss & Horn [1983] and Jain [1983].

A lot of the published work has been based on what we have called the discrete approach. In this context, it is important to establish the minimum number of points that force a unique solution. In the process, much of what is known in photogrammetry, discussed earlier in chapter 13, has been rediscovered.

Clocksin [1980] attempted to recover the motion from derivatives of the optical flow. Neumann [1980b] showed that changes in shading due to motion do not constitute a solid basis for recovering motion information. Prazdny [1980, 1981] tackled the passive navigation and structure-from-motion problems directly. This earlier work is refined in Prazdny [1983]. Nagel [1981b] and Dreschler & Nagel [1982] worked on recovering the shape and motion of a moving object and explored the representational issues. Jain [1981] wrote another early paper on the recovery of motion and depth from optical flow. Ballard & Kimball [1983] tried to determine rigid body motion from optical flow. Webb & Aggarwal [1982] extended the analysis to multiple objects mutually mechanically restrained.

Tsai, Huang, & Zhu [1982] started the careful analysis of optical flow in the case of a planar patch. Hay [1966] apparently was the first to mention the two-way ambiguity noted by Tsai, Huang, & Zhu. Waxman & Ullman [1983] also addressed the two-way ambiguity for planar surfaces. For shorter proofs of the main result see Maybank [1984] and Negahdaripour & Horn [1985].

Bruss & Horn [1983] tackled the problem using the least-squares method advocated in this chapter. Fang & Huang [1984a,b], Jerian & Jain [1984], and Sugihara & Sugie [1984] still used the discrete approach, however. Tsai & Huang [1984a,b] tried to show that the motion can be recovered uniquely when the object's surface is curved and dealt with the case in which three successive images are available. Waxman & Wohn [1984] dealt with the planar surface case in more detail.

Other important papers include those by Koenderink & van Doorn [1976], Neumann [1980a], Longuet-Higgins [1981], Tsai [1982], Hoffman & Flinchbaugh [1982], Yen & Huang [1983], and Adiv [1984].

Negahdaripour & Horn [1985] found a way to go directly from brightness gradients to the motion of the observer without computing the optical flow or matching discrete features. Using a simplified notation, they found a shorter proof of the result that there are two planar surfaces giving rise to the same instantaneous motion field.

17.7 Exercises

17-1 Here we extend some of the results developed in chapter 12, where we discussed optical flow, using the equations developed for the translational and rotational components.

(a) Show that the Laplacian of the optical flow is zero when a camera viewing a planar surface is translating parallel to the image plane. The plane being viewed need not be parallel to the image plane.

(b) Show that the Laplacian of the optical flow is zero when a camera is translating relative to a plane that is orthogonal to the optical axis. The translation need not be parallel to the image plane.

(c) Show that the Laplacian of the optical flow is not zero when the camera is rotating about an axis other than the optical axis.

17-2 In the case of pure translation, the optical flow consists of vectors of varying length that all pass through a single point when extended. This point where the optical flow is zero is called the *focus of expansion*; it is the image of the ray along which the camera moves. How are the coordinates of the focus of expansion related to the parameters U, V, and W? Show that in the case of pure translation the motion can be determined (up to a scale factor) by considering the direction of the optical flow at two image points. Explain why this may not be the best way to use optical flow information.

17-3 In the case of pure rotation, the optical flow consists of vectors whose lengths do not depend on the distances of the objects. Instead, the magnitudes and directions of the vectors are determined by the axis of rotation and the angular velocity. Define the *center of rotation* to be the point where the optical flow is zero, somewhat analogous to the focus of expansion in the previous problem. Show that in the case of pure rotation, the motion can be determined from the optical flow at two points. Demonstrate further that at one of these points we need to know only the direction of the flow. Explain why this may not be the best way to use optical flow information.

17-4 An alternate approach to passive navigation involves the identification and tracking of "features." Using the methods developed in our discussion of photogrammetry and stereo, determine the smallest number of points on an object that fully determine its motion when they are identified in two images taken a small interval apart. Hint: This is nontrivial.

17-5 Show that if $f(k\mathbf{r}) = f(\mathbf{r})$, then $\nabla f(\mathbf{r}) \cdot \mathbf{r} = 0$, where ∇ produces the gradient, a vector whose components are the derivatives of f with respect to x, y, and z.

17-6 Suppose that $g(\mathbf{t}) = \mathbf{t}^T \mathbf{G} \mathbf{t}$. Show that the extrema of $g(\mathbf{t})$ correspond to the eigenvectors of the matrix $\mathbf{G}$ if $\mathbf{t}$ is constrained to be a unit vector. Hint: Introduce a Lagrange multiplier, as shown in the appendix.

17-7 Consider the case of purely translational motion. Show that minimization with respect to the ML_{Zuv} norm leads to the same equations as minimization with respect to the $ML_{\alpha\beta}$ norm. Explain why this should be so.

17-8 Consider the case of purely translational motion. Suppose we carry out the minimization with respect to the ML_Z norm defined by

$$\|f(x,y)\|_Z = \iint \left(f(x,y)Z(x,y)\right)^2 dx\,dy.$$

Here optical flow velocities for points that are farther away are weighted more heavily. Show that the result in this case depends only on the direction of the optical flow at each point, not on its magnitude.

17-9 Show that in the case of a rectangular image plane of width $2W$ and height $2H$,

$$\overline{a} = 4WH\left(\frac{H^4}{5} + \frac{2H^2}{3} + 1\right) + \frac{4W^3H^3}{9},$$

$$\overline{b} = 4WH\left(\frac{W^4}{5} + \frac{2W^2}{3} + 1\right) + \frac{4W^3H^3}{9},$$

$$\overline{c} = \frac{4WH}{3}\left(W^2 + H^2\right),$$

$$\overline{d} = \overline{e} = \overline{f} = 0,$$

where $\overline{a}$, $\overline{b}$, $\overline{c}$, $\overline{d}$, $\overline{e}$, and $\overline{f}$ are the integrals defined above for the purely rotational case. Show that the matrix $\mathbf{M}$ is not singular and find its inverse.

17-10 Under somewhat impoverished conditions, it is possible for a passive navigation problem to have more than one solution. Here we explore the case in which a planar surface is being imaged.

(a) Suppose that the surface being imaged is the plane and that

$$Z = Z_0 + pX + qY,$$

where X, Y, and Z are coordinates of points on the surface. Show that

$$\frac{Z_0}{Z} = 1 - px - qy,$$

where x and y are image coordinates.

(b) Show that the motion field in this case is given by the following second-order polynomials in x and y:

$$u = u_t + u_r = \frac{1}{Z_0}(-U + xW)(1 - px - qy) + Axy - B(x^2 + 1) + Cy,$$

$$v = v_t + v_r = \frac{1}{Z_0}(-V + yW)(1 - px - qy) + A(y^2 + 1) - Bxy - Cx.$$