CHAPTER

17 Invariants and Perspective

Invariants are important for achieving recognition in both 2-D and 3-D. The basic idea is to identify any parameters that do not vary between different instances of the same object. Unfortunately, perspective projection makes the issue far harder in the general 3-D case. This chapter explores the problem and demonstrates a number of useful techniques. At the same time it explores the problems of perspective projection, with some interesting outcomes.

Look out for:

- how a ratio of distances between features along the same straight line can act as a convenient invariant under weak perspective projection.
- how a ratio of ratios (or "cross-ratio") can act as a convenient invariant under full perspective projection.
- how the cross-ratio type of invariant can rather cunningly be generalized to cover many wider possibilities.
- how the cross-ratio type of invariant seems largely unable to provide invariance outside any given plane.
- vanishing point detection and its relevance to image interpretation.
- use of invariance for initiating face recognition.
- how to optimize views of 2-D pictures to limit perspective distortions.
- problems involved in "stitching" photographs.

While this chapter considers only one aspect of 3-D vision, it is extremely useful both in helping to cue into complex images (see particularly the egomotion example in Fig. 17.4 and the facial analysis example in Section 17.10) and in taking shortcuts around the tedious analysis of 3-D geometry (see, e.g., Sections 17.8 and 17.9).

17.1 INTRODUCTION

Pattern recognition is a complex task and, as stated in Chapter 1, involves the twin processes of discrimination and generalization. Indeed, the latter process is in many ways more important than the first—especially in the initial stages of recognition—since there is so much redundant information in a typical image. Thus, we need to find ways of helping to eliminate invalid matches. This is where the study of invariants comes into its own.

An invariant is a property of an object or class of objects that does not change with changes of viewpoint or object pose and which can therefore be used to help distinguish it from other objects. The procedure is to search for objects with a specific invariant, so that those which do not possess the invariant can immediately be discarded from consideration. An invariant property can be regarded as a necessary condition for an object to be in the chosen class, although in principle only detailed subsequent analysis will confirm the presence of that object in the chosen class. In addition, if an object is found to possess the correct invariant, it will then be profitable to pursue the analysis further and find its pose, size, or other relevant data. Ideally an invariant would uniquely identify an object as being of a particular type or class. Thus, an invariant should not merely be a property that leads to further hypotheses being made about the object, but one that fully characterizes it. However, the difference is a subtle one, more a matter of degree and purpose than an absolute criterion. We shall see later in this chapter the extent of the difference by appealing to a number of specific cases.

Let us first consider an object being viewed from directly overhead at a known distance by a camera whose optical axis is normal to the plane on which the object is lying. We shall assume that the object is flat. Take two-point features on the object, such as corners or small holes. If we measure the image distance between these features, this acts as an invariant, in that:

1. it has a value independent of the translation and orientation parameters of the object;
2. it will be unchanged for different objects of the same type; and
3. it will in general be different from the distance parameters of other objects that might be on the object plane.

Thus, measurement of distance provides a certain lookup or indexing quality that ideally identifies the object uniquely, although further analysis will be required to fully locate it and ascertain its orientation. Hence, distance has all the requirements of an invariant, although it could also be argued that it is only a feature that *helps* to classify objects. Clearly, we are here ignoring an important factor—the effect of imprecision in measurement—due to spatial quantization (or inadequate spatial resolution), noise, lens distortions, and so on; in addition, the effects of partial occlusion or breakage are also being ignored. Most definitely, there is a limit to what can be achieved with a single invariant measure, although

in what follows we attempt to reveal what is possible and demonstrate the advantages of employing an invariant-orientated approach.

The above ideas relating to distance as an invariant measure showed it to be useful in suppressing the effects of translations and rotations of objects in 2-D. Hence, it is of little direct value when considering translations and rotations in 3-D. Furthermore, it is not even able to cope with scale variations of objects in 2-D. Moving the camera closer to the object plane and refocussing totally change the situation, and all values of the distance invariant residing in the object indexing table must be changed and the old values ignored. However, a little thought will show that this last problem could be overcome. All we need to do is to take *ratios* of distances. This requires a minimum of three point features to be identified in the image and the inter-feature distances measured. If we call two of these distances d_1 and d_2, then the ratio d_1/d_2 will act as a scale-independent invariant, i.e., we will be able to identify objects using a single indexing operation whatever their 2-D translation, orientation, or apparent size or scale. An alternative to this idea is to measure the angle between pairs of distance vectors, $\cos^{-1}(\mathbf{d}_1.\mathbf{d}_2/|d_1||d_2|)$, which will again be scale invariant.

Of course, this consideration has already been invoked in our earlier work on shape analysis. If objects are subject only to 2-D translations and rotations but not to changes of scale, they can be characterized by their perimeters or areas as well as their normal linear dimensions; furthermore, parameters such as compactness and aspect ratio, which employ dimensionless ratios of image measurements, have been acknowledged in Chapter 9 to overcome the size/scale problem.

Nevertheless, the main motivation for using invariants is to obtain mathematical measures of configurations of object features that are carefully designed to be independent of the viewpoint or coordinate system used, and indeed to not require specific setup or calibration of the image acquisition system. However, it must be emphasized that camera distortions are assumed to be absent or to have been compensated for by suitable post-camera transformations (see Chapter 18).

This chapter proceeds to develop the above ideas and later applies them to vanishing point detection (Sections 17.7–17.9), face recognition (Section 17.10), obtaining optimal views of 2-D pictures, and the "stitching" of digital photographs (Section 17.11). Interestingly, with the small intellectual outlay of the initial sections, these applications emerge with very little additional effort demonstrating the value of the basic theory presented here.

17.2 CROSS-RATIOS: THE "RATIO OF RATIOS" CONCEPT

It would be most useful if we could extend the above ideas to permit indexing for general transformations in 3-D. Indeed, an obvious question is whether finding ratios of ratios of distances will provide suitable invariants and lead to such a generalization. The answer is that ratios of ratios do provide useful further invariants,

although going further than this leads to considerable complication, and there are restrictions on what can be achieved with limited computation. In addition, noise ultimately becomes a limiting factor, since so many parameters become involved in the computation of complex invariants that the method ultimately loses steam (it becomes just one of many ways of raising hypotheses and therefore has to compete with other approaches in a manner appropriate to the particular problem application being studied).

We now consider the ratio of ratios approach. Initially, we only examine a set of four collinear points on an object. Figure 17.1 shows such a set of four points (P_1, P_2, P_3, P_4) and a transformation of them (Q_1, Q_2, Q_3, Q_4) such as that produced by an imaging system with optical center C (c, d). Choice of a suitable pair of oblique axes permits the coordinates of the points in the separate sets to be expressed respectively as:

$$(x_1, 0), (x_2, 0), (x_3, 0), (x_4, 0)$$
$$(0, y_1), (0, y_2), (0, y_3), (0, y_4)$$

Taking points P_i, Q_i, we can write the ratio CQ_i:PQ_i both as $c/(-x_i)$ and as $(d - y_i)/y_i$. Hence:

$$\frac{c}{x_i} + \frac{d}{y_i} = 1 \tag{17.1}$$

which must be valid for all i. Subtraction of the ith and jth relations now gives:

$$\frac{c(x_j - x_i)}{x_i x_j} = \frac{-d(y_j - y_i)}{y_i y_j} \tag{17.2}$$

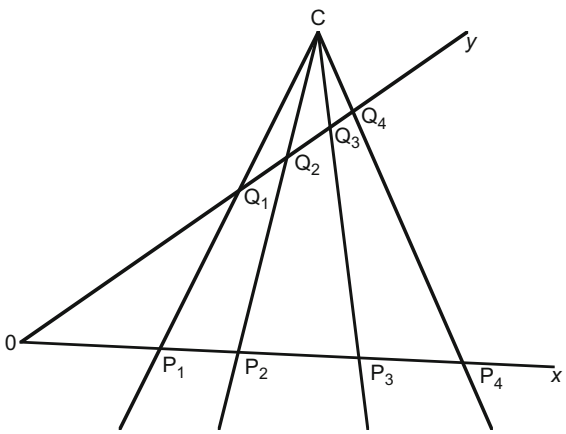

FIGURE 17.1

Perspective transformation of four collinear points. This figure shows four collinear points (P_1, P_2, P_3, P_4) and a transformation of them (Q_1, Q_2, Q_3, Q_4) similar to that produced by an imaging system with optical center C. Such a transformation is called a *perspective transformation*.

Forming a ratio between two such relations will now eliminate the unknowns c and d. For example, we will have:

$$\frac{x_3(x_2 - x_1)}{x_2(x_3 - x_1)} = \frac{y_3(y_2 - y_1)}{y_2(y_3 - y_1)} \tag{17.3}$$

However, the result still contains factors such as x_3/x_2 that depend on absolute position. Hence, it is necessary to form a suitable ratio of such results, which cancels out the effects of absolute positions:

$$\frac{(x_2 - x_4)/(x_3 - x_4)}{(x_2 - x_1)/(x_3 - x_1)} = \frac{(y_2 - y_4)/(y_3 - y_4)}{(y_2 - y_1)/(y_3 - y_1)} \tag{17.4}$$

Thus, our original intuition that a ratio of ratios type of invariant might exist that would cancel out the effects of a perspective transformation is correct. In particular, four collinear points viewed from any perspective viewpoint yield the same value of the cross-ratio, defined as above. The value of the cross-ratio of the four points is written:

$$C(\mathrm{P}_1, \mathrm{P}_2, \mathrm{P}_3, \mathrm{P}_4) = \frac{(x_3 - x_1)(x_2 - x_4)}{(x_2 - x_1)(x_3 - x_4)} \tag{17.5}$$

For clarity, we shall write this particular cross-ratio as κ in what follows. Note that there are $4! = 24$ possible ways in which 4 collinear points can be ordered on a straight line, and hence there could be 24 cross-ratios. However, they are not all distinct, and in fact there are only six different values. To verify this we start by interchanging pairs of points:

$$C(\mathrm{P}_2, \mathrm{P}_1, \mathrm{P}_3, \mathrm{P}_4) = \frac{(x_3 - x_2)(x_1 - x_4)}{(x_1 - x_2)(x_3 - x_4)} = 1 - \kappa \tag{17.6}$$

$$C(\mathrm{P}_1, \mathrm{P}_3, \mathrm{P}_2, \mathrm{P}_4) = \frac{(x_2 - x_1)(x_3 - x_4)}{(x_3 - x_1)(x_2 - x_4)} = \frac{1}{\kappa} \tag{17.7}$$

$$C(\mathrm{P}_1, \mathrm{P}_2, \mathrm{P}_4, \mathrm{P}_3) = \frac{(x_4 - x_1)(x_2 - x_3)}{(x_2 - x_1)(x_4 - x_3)} = 1 - \kappa \tag{17.8}$$

$$C(\mathrm{P}_4, \mathrm{P}_2, \mathrm{P}_3, \mathrm{P}_1) = \frac{(x_3 - x_4)(x_2 - x_1)}{(x_2 - x_4)(x_3 - x_1)} = \frac{1}{\kappa} \tag{17.9}$$

$$C(\mathrm{P}_3, \mathrm{P}_2, \mathrm{P}_1, \mathrm{P}_4) = \frac{(x_1 - x_3)(x_2 - x_4)}{(x_2 - x_3)(x_1 - x_4)} = \frac{\kappa}{\kappa - 1} \tag{17.10}$$

$$C(\mathrm{P}_1, \mathrm{P}_4, \mathrm{P}_3, \mathrm{P}_2) = \frac{(x_3 - x_1)(x_4 - x_2)}{(x_4 - x_1)(x_3 - x_2)} = \frac{\kappa}{\kappa - 1} \tag{17.11}$$

These cases provide the main possibilities, but of course interchanging more points will yield a limited number of further values—in particular:

$$C(\mathrm{P}_3, \mathrm{P}_1, \mathrm{P}_2, \mathrm{P}_4) = 1 - C(\mathrm{P}_1, \mathrm{P}_3, \mathrm{P}_2, \mathrm{P}_4) = 1 - \frac{1}{\kappa} = \frac{\kappa - 1}{\kappa} \tag{17.12}$$

$$C(\mathrm{P}_2, \mathrm{P}_3, \mathrm{P}_1, \mathrm{P}_4) = \frac{1}{C(\mathrm{P}_2, \mathrm{P}_1, \mathrm{P}_3, \mathrm{P}_4)} = \frac{1}{1-\kappa} \tag{17.13}$$

This covers all six cases, and a little thought (based on trying further interchanges of points) will show there can be no others (we can only repeat κ, $1-\kappa$, $\kappa/(\kappa-1)$, and their inverses). Of particular interest is the fact that numbering the points in reverse (which would correspond to viewing the line from the other side) leaves the cross-ratio unchanged. Nevertheless, it is inconvenient that the same invariant has six different manifestations, as this implies that six different index values have to be looked up before the class of an object can be ascertained. On the other hand, if points are labeled in order along the line rather than randomly, it should generally be possible to circumvent this situation.

So far we have been able to produce only one projective invariant, and this corresponds to the rather simple case of four collinear points. The usefulness of this measure is augmented considerably when it is noted that four collinear points, taken in conjunction with another point, define a pencil[1] of concurrent coplanar lines passing through the latter point. Clearly, we can assign a unique cross-ratio to this pencil of lines, equal to the cross-ratio of the collinear points on any line passing through them. We can clarify the situation by considering the angles between the various lines (Fig. 17.2). Applying the sine rule four times to determine the four distances in the cross-ratio $C(\mathrm{P}_1, \mathrm{P}_2, \mathrm{P}_3, \mathrm{P}_4)$ gives:

$$\frac{x_3 - x_1}{\sin\alpha_{13}} = \frac{\mathrm{OP}_1}{\sin\beta_3} \tag{17.14}$$

$$\frac{x_2 - x_4}{\sin\alpha_{24}} = \frac{\mathrm{OP}_4}{\sin\beta_2} \tag{17.15}$$

$$\frac{x_2 - x_1}{\sin\alpha_{12}} = \frac{\mathrm{OP}_1}{\sin\beta_2} \tag{17.16}$$

$$\frac{x_3 - x_4}{\sin\alpha_{34}} = \frac{\mathrm{OP}_4}{\sin\beta_3} \tag{17.17}$$

Substituting in the cross-ratio formula (Eq. (17.5)) and canceling the factors OP_1, OP_4, $\sin\beta_2$, and $\sin\beta_3$ now give:

$$C(\mathrm{P}_1, \mathrm{P}_2, \mathrm{P}_3, \mathrm{P}_4) = \frac{\sin\alpha_{13}\,\sin\alpha_{24}}{\sin\alpha_{12}\,\sin\alpha_{34}} \tag{17.18}$$

Thus, the cross-ratio depends only on the angles of the pencil of lines. It is interesting that appropriate juxtaposition of the sines of the angles gives the final

[1] It is a common nomenclature of projective geometry to call a set of concurrent lines a *pencil* (e.g., Semple and Kneebone, 1952).

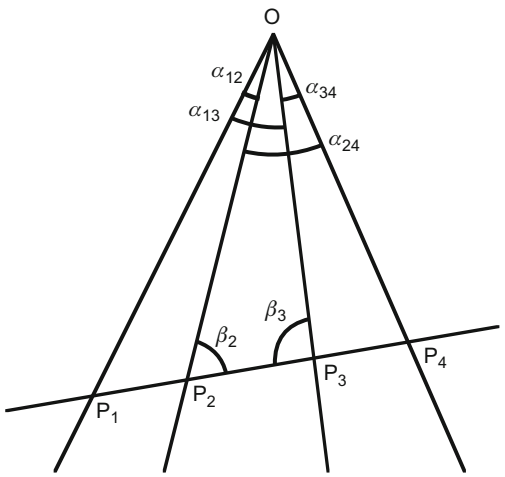

FIGURE 17.2

Geometry for calculation of the cross-ratio of a pencil of lines. The figure shows the geometry required to calculate the cross-ratio of a pencil of lines in terms of the angles between them.

formula for invariance under perspective projection: using the angles themselves would not give the desired degree of mathematical invariance. Indeed, we can immediately see one reason for this: inversion of the direction of any line must leave the situation unchanged, so the formula must be tolerant to adding π to each of the two angles linking the line; this could not be achieved if the angles appeared without suitable trigonometric functions.

We can extend this concept to four concurrent planes since the concurrent lines can be projected into four concurrent planes once a separate axis for the concurrency has been defined. As there are infinitely many such axes, there are infinitely many ways in which sets of planes can be chosen. Thus, the original simple result on collinear points can be extended to a much more general case.

Finally, note that we started by trying to generalize the case of four collinear points, but what we achieved was first to find a dual situation in which points become lines also described by a cross-ratio, and then to find an extension in which planes are described by a cross-ratio. We now return to the case of four collinear points and see how we can extend it in other ways.

17.3 INVARIANTS FOR NONCOLLINEAR POINTS

First, imagine that not all the points are collinear: specifically, let us assume that one point is not in the line of the other three. If this is the case, there is not

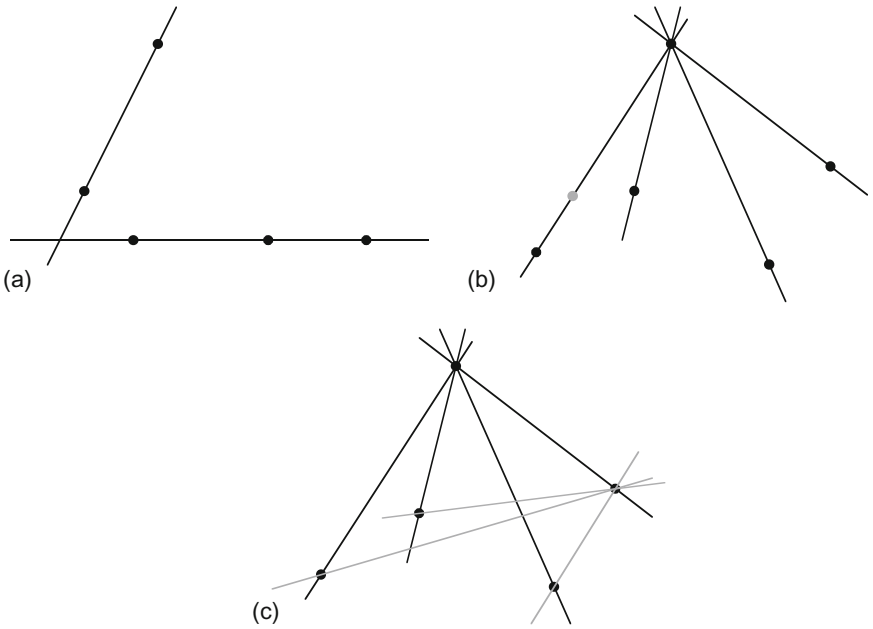

FIGURE 17.3

Calculation of invariants for a set of noncollinear points. Part (a) shows how the addition of a fifth point to a set of four points, one of which is not collinear with the rest, permits the cross-ratio to be calculated. Part (b) shows how the calculation can be extended to any set of noncollinear points; also shown is an additional (gray) point that a single cross-ratio fails to distinguish from other points on the same line. Part (c) shows how any failure to identify a point uniquely can be overcome by calculating the cross-ratio of a second pencil generated from the five original points.

enough information to calculate a cross-ratio. However, if a further coplanar point is available, we can draw an imaginary line between the noncollinear points to intersect their common line in a unique point, which will then permit a cross-ratio to be computed (Fig. 17.3(a)). Nevertheless, this is some way from a general solution to the characterization of a set of noncollinear points. We might inquire how many point features in general position[2] on a plane will be required to calculate an invariant. In fact, the answer is five, since the fact that we can form a cross-ratio from the angles between four lines immediately means that forming a pencil of four lines from five points defines a cross-ratio invariant (Fig. 17.3(b)).

[2]Points on a plane that are chosen at random, and that are not collinear or in any special pattern such as a regular polygon, are described as being *in general position*.

While the value of this cross-ratio provides a necessary condition for a match between two sets of five general coplanar points, it could be a fortuitous match because the condition depends only on the relative directions between the various points and the reference point, i.e., any of the nonreference points is only defined to the extent that it lies on a given line. Clearly, two cross-ratios formed by taking two reference points will define the directions of all the remaining points uniquely (Fig. 17.3(c)).

We can now summarize the general result, which stipulates that for five general coplanar points, no three being collinear, two different cross-ratios are required to characterize the shape. These cross-ratios correspond to taking in turn two separate points and producing pencils of lines passing through them and (in each case) the remaining four points (Fig. 17.3(c)). While it might appear that at least five cross-ratios result from this sort of procedure, there are only two functionally independent cross-ratios—essentially because the position of any point is defined once its direction relative to two other points is known.

Next, we consider the problem of finding the ground plane in practical situations—especially that of egomotion including vehicle guidance (Fig. 17.4). Here a set of four collinear points can be observed from one frame to the next. If they are on a single plane, then the cross-ratio will remain constant, but if one is elevated above the ground plane (as, e.g., a bridge or another vehicle), then the cross-ratio will vary over time. Taking a larger number of points, it should clearly be possible to deduce which are on the ground plane and which are not, by using a process of elimination; however, the amount of noise and clutter will determine the computational complexity of the task. Note that this is possible without any calibration of the camera, this being perhaps the main value of concentrating attention on projective invariants. Note also that there is a potential problem regarding irrelevant planes, such as the vertical faces of buildings. The cross-ratio test is so resistant to viewpoint and pose that it merely ascertains whether the points being tested are coplanar. It is only by using a sufficiently large number of independent sets of points that one plane can be discriminated from another (for simplicity we ignore here any subsequent stages of pose analysis that might be carried out).

17.3.1 Further Remarks About the Five-Point Configuration

The above description outlines the principles for solving the five-point invariance problem, but does not show clearly the conditions under which it is guaranteed to operate properly. In fact, these are straightforward to demonstrate. First, the cross-ratio can be expressed in terms of the sines of the angles α_{13}, α_{24}, α_{12}, α_{34}. Next, these can be re-expressed in terms of areas of relevant triangles, using equations typified by the following to express area:

$$\Delta_{513} = \frac{1}{2} a_{51} a_{53} \sin \alpha_{13} \tag{17.19}$$

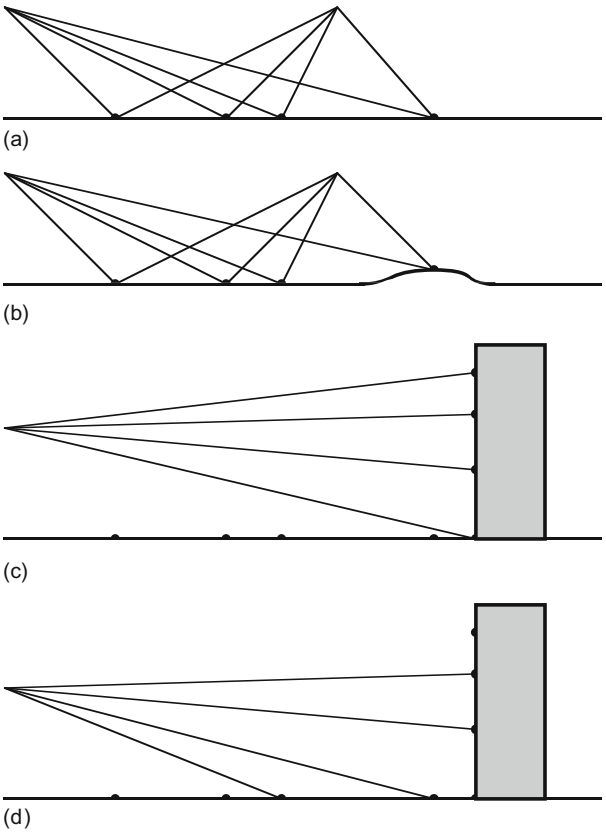

FIGURE 17.4

Use of cross-ratio for egomotion guidance. Part (a) shows how the cross-ratio for a set of four collinear points can be tracked to confirm that the points are collinear. This suggests that they lie on the ground plane. Part (b) shows a case where the cross-ratio will not be constant. Part (c) shows a case where the cross-ratio is constant, although they actually lie on a plane that is not the ground plane. Part (d) shows a case where all four points lie on planes, yet the cross-ratio will not be constant.

Finally, the area can be re-expressed in terms of the point coordinates in the following way:

$$\Delta_{513} = \frac{1}{2}\begin{vmatrix} p_{5x} & p_{1x} & p_{3x} \\ p_{5y} & p_{1y} & p_{3y} \\ p_{5z} & p_{1z} & p_{3z} \end{vmatrix} = \frac{1}{2}|\mathbf{p}_5\ \mathbf{p}_1\ \mathbf{p}_3| \tag{17.20}$$

Using this notation, a suitable final pair of cross-ratio invariants for the configuration of five points may be written:

$$C_a = \frac{\Delta_{513}\Delta_{524}}{\Delta_{512}\Delta_{534}} \tag{17.21}$$

$$C_b = \frac{\Delta_{124}\Delta_{135}}{\Delta_{123}\Delta_{145}} \tag{17.22}$$

While three more such equations may be written down, these will not be independent of the other two and will not carry any further useful information.

Note that a determinant will go to zero or infinity if the three points it relates to are collinear, corresponding to the situation when the area of the triangle is zero. Clearly, when this happens, any cross-ratio containing this determinant will no longer be able to pass on any useful information. On the other hand, there is actually no further information to pass on, as this now constitutes a special case that is describable by a single cross-ratio: we have reverted to the situation shown in Fig. 17.3(a).

Finally, Fig. 17.3 misses out one further interesting case: the situation of two points and two lines (Fig. 17.5). Constructing a line joining the two points and producing it until it meets the two lines, we then have four points on a single line; thus, the configuration is characterized by a single cross-ratio. Note also that the two lines can be extended until they intersect, and further lines can be constructed from the intersection to meet the two points. This gives a pencil of lines characterized by a single cross-ratio (Fig. 17.5(c)); the latter must have the same value as that computed for the four collinear points.

17.4 INVARIANTS FOR POINTS ON CONICS

These discussions clearly help to build up an understanding of how geometric invariants can be designed to cope with sets of points, lines, and planes in 3-D. Significantly more difficult is the case of curved lines and surfaces, although much headway has now been made with regard to the understanding of conics and certain other surfaces (see Mundy and Zisserman, 1992a). It will not be possible to examine all such cases in depth here. However, it will be useful to consider conic sections and particularly ellipses in more detail.

First, we consider Chasles' theorem, which dates from the 19th century. (The history of projective geometry is quite rich and was initially carried out totally independently of the requirements of machine vision.) Suppose we have four fixed coplanar points F_1, F_2, F_3, F_4 on a conic section curve and one variable point P in the same plane (Fig. 17.6). Then the four lines joining P to the fixed points form a

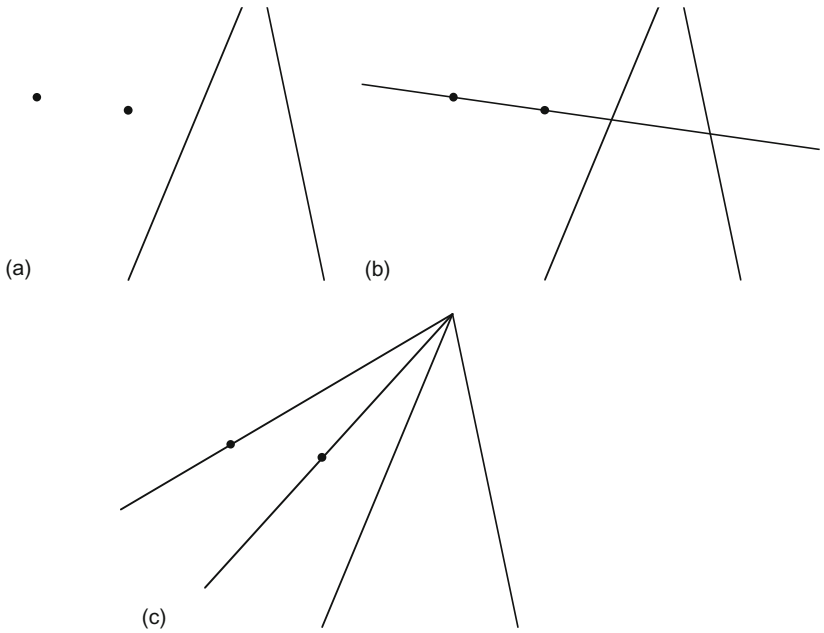

FIGURE 17.5

Cross-ratio for two lines and two points. (a) Basic configuration. (b) How the line joining the two points introduces four collinear points to which a cross-ratio may be applied. (c) How joining the two points to the junction of the two lines creates a pencil of four lines to which a cross-ratio may be applied.

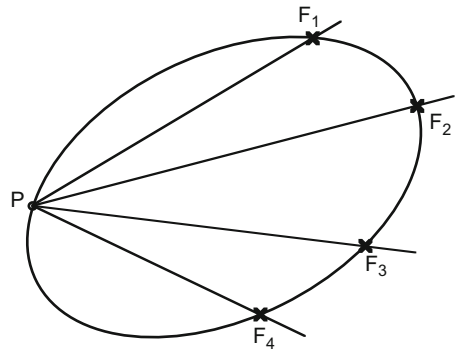

FIGURE 17.6

Definition of a conic using a cross-ratio. Here P is constrained to move so that the cross-ratio of the pencil from P to F_1, F_2, F_3, F_4 remains constant. By Chasles' theorem, P traces out a conic curve.

pencil whose cross-ratio will in general vary with the position of P. Chasles' theorem states that if P now moves so as to keep the cross-ratio constant, then P will trace out a conic section. This clearly provides a means of checking whether a set of points lies on a planar curve such as an ellipse. Note the close analogy with the problem of ground plane detection already mentioned. Again the amount of computation could become excessive if there were a lot of noise or clutter in the image. When the image contains N boundary features that need to be checked out, the problem complexity is intrinsically $O(N^5)$, since there are $O(N^4)$ ways of selecting the first four points, and for each such selection, $N-4$ points must be examined to determine whether they lie on the same conic. However, choice of suitable heuristics would be expected to limit the computation. Note the problem of ensuring that the first four points are tested in the same order around the ellipse, which is liable to be tedious (a) for point features and (b) for disconnected boundary features.

While Chasles' theorem gives an excellent opportunity to use invariants to locate conics in images, it is not at all discriminatory in this. The theorem applies to a general conic, hence it does not immediately permit circles, ellipses, parabolas, or hyperbolas to be distinguished, a fact that would sometimes be a distinct disadvantage. This is an example of a more general problem in pattern recognition system design—of deciding exactly how and in what sequence one object should be differentiated from another. Because of the space constraint, this point is not considered further here.

Finally, we state without proof that conic section curves can all be transformed under perspective projection to other types of conic section, and thus into ellipses; subsequently they can be transformed into circles. Thus, any conic section curve can be transformed projectively into a circle, while the inverse transformation can transform it back again (Mundy and Zisserman, 1992b). This means that simple properties of the circle can frequently be generalized to ellipses or other conic sections. In this context, points to bear in mind are that, after perspective projection, lines intersecting curves do so in the same number of points, and thus tangents transform into tangents and chords into chords, three-point contact (in the case of non-conic curves) remains three-point contact, and so on. Returning to Chasles' theorem, a simple proof in the case of circles will automatically generalize to more complex conic section curves.

In response to this assertion, we can in fact derive Chasles' theorem almost trivially for a circle. Appealing to Fig. 17.7, we see that the angles φ_1, φ_2, φ_3 are equal to the respective angles γ_1, γ_2, γ_3 (angles in the same segment of a circle). Thus, the pencils PF_1, PF_2, PF_3, PF_4 and QF_1, QF_2, QF_3, QF_4 have equal angles, their *relative* directions being superposable. This means that they will have the same cross-ratio, defined by Eq. (17.18). Hence, the cross-ratio of the pencil will remain constant as P traces out the circle. As stated above, the property will automatically generalize to any other conic.

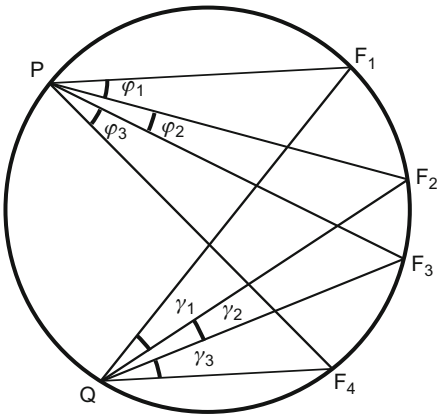

FIGURE 17.7

Proof of Chasles' theorem. This diagram shows that the four points F_1, F_2, F_3, F_4 subtend the same angles at P as they do at the fixed point Q. Thus, the cross-ratio is the same for all points on the circle. This means that Chasles' theorem is valid for a circle.

17.5 DIFFERENTIAL AND SEMI-DIFFERENTIAL INVARIANTS

There have been many attempts to characterize continuous curves by invariants. The obvious way forward is to represent points on a curve in terms of local curve derivatives. If a sufficient number of these can be obtained, invariants can be formed and computed. However, the noise (including digitization noise) that always exists on curves limits the accuracy of higher derivatives, and as a result it is difficult to form useful invariants in this way. In general, the second derivative of the curve function is the highest that can normally be used, and this corresponds to curvature, which is only an invariant for Euclidean transformations (translation and rotation without change of scale).

As a result of this problem, semi-differential invariants are often used instead of differential invariants. They involve considering only a few "distinguished" points on curves, and using these to generate invariants. The most common distinguished points to be used in this way are (Fig. 17.8) the following:

1. Points of inflection
2. Sharp corners on curves
3. Cusps on curves (corners where the bounding tangents are coincident)
4. Bitangent points (points of contact of a line that touches the curve twice)
5. Other points whose locations can be derived from existing distinguished points by geometric constructions

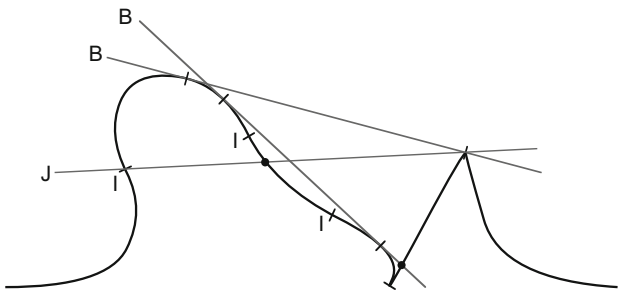

FIGURE 17.8

Means for finding distinguished points on a curve. The two bitangents contact the curve in a total of four bitangent points. Three points of inflection I provide another three distinguished points. A cusp and a corner provide a further two distinguished points (the latter also being a bitangent point). The line marked J contributes a further distinguished point on the curve, as does one of the bitangents: these are marked as large dots rather than as short lines.

Tangent points are unlikely to be suitable, and hence are not included in this list, as a smooth curve will have tangents along its entire length. This is because they are characterized merely by two-point contact between a limiting chord and the curve. However, a point of inflection represents a three-point contact, and this means that it will be reasonably well localized, and its tangent will have a well-defined direction. On the other hand, bitangent points will be even more accurately represented, as the tangent direction will be accurately defined by two well-separated points on the curve (Fig. 17.8): nevertheless bitangent points will still incur some longitudinal error.

Bitangents can be of several sorts: in particular, they can contact the same shape on the same side; they can also cross the body and contact it on both sides. This latter case is more complex and is therefore sometimes discounted in machine vision applications. Nevertheless, it provides a means of finding further invariant reference points on an object. Note that this clearly happens directly, in that the bitangent points are already distinguished points. It also happens indirectly, as the bitangent may cross other reference lines, thereby defining further distinguished points. Figure 17.9 shows several cases of direct and indirect distinguished points, the most accurate of which arise from bitangents, while slightly less accurate ones arise from points of inflection.

Once enough distinguished points and reference lines between them have been found, cross-ratio invariants may be obtained (a) from the incidence of distinguished points lying along suitable reference lines and (b) from pencils of lines drawn from distinguished points to line crossings or to other distinguished points.

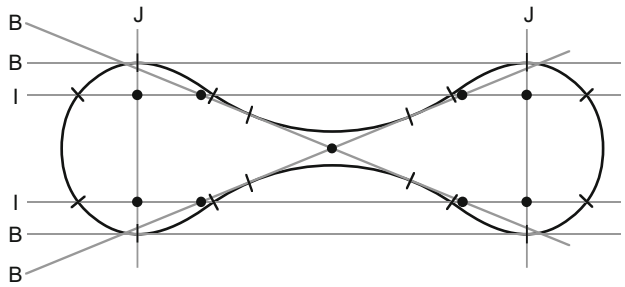

FIGURE 17.9

Means for finding direct and indirect distinguished points for an object. The four lines marked B are bitangents, which contribute twelve bitangent points: two of the bitangents contact the object on opposite sides of its boundary. The two lines marked I arise from points of inflection. The two lines marked J are joins of bitangent points. The nine large dots are indirect distinguished points, which do not lie on the object boundary. Clearly, a good many more indirect distinguished points could be generated, although not all would have accurately defined locations.

A remark is needed to confirm that points of inflection can act as suitable distinguished points that are invariant under perspective transformations. Starting from the premise that perspective transformations preserve straight lines and points arising from crossings between curves and lines, we note that a chord that crosses a curve three times will also cross it three times under perspective projection. This will still apply even when the three crossing points merge into three-point contact.[3] Hence, points of inflection are suitable distinguished points and are perspective-invariant.

This treatment has only dealt with planar curves, and has not covered spatial nonplanar curves. The latter is a significantly more difficult area, as concepts such as bitangents and points of inflection have to be assigned new meaning in this more general domain. It is a subject we shall not be able to broach here.

17.6 SYMMETRIC CROSS-RATIO FUNCTIONS

When applying a cross-ratio to a set of points on a line, it frequently happens that the order of the points on the line is known. For example, this will almost certainly be the case if feature detection of an image is carried out in a forward raster scan. Hence, the only confusion in the ordering will be the direction in which the

[3]Three-point contact is distinguishable from two-point contact in that the tangent crosses the curve at the point of contact.

line has been traversed. However, the cross-ratio is independent of the end from which the line is scanned, since $C(P_1, P_2, P_3, P_4) = C(P_4, P_3, P_2, P_1)$. Nevertheless, there are situations in which the ordering of the cross-ratio features will not be known with certainty. This may occur for the situations shown in Figs. 17.3 and 17.5, where the features themselves do not all lie on a single line, where the features are angles, or where the points lie on a conic whose equation is as yet unknown. In such circumstances, it will be useful to have an invariant that takes in all possible orders of the features.

To derive such an invariant, note first that if there is a confusion in the ordering of the points such that the value could be either κ or $(1 - \kappa)$, then we could apply the function $f(\kappa) = \kappa(1 - \kappa)$, which has the property $f(\kappa) = f(1 - \kappa)$, and this will solve the problem. Alternatively, if there is confusion between κ and $1/\kappa$, then we could apply the function $g(\kappa) = \kappa + 1/\kappa$, which has the property $g(\kappa) = g(1/\kappa)$, and again this will solve the problem.

However, if there is potential confusion between the values κ, $(1 - \kappa)$, and $1/\kappa$, the situation becomes more complicated. It is difficult to write down any obvious function that satisfies the double condition $h(\kappa) = h(1 - \kappa) = h(1/\kappa)$, although we may have a soundly based intuition that it will involve symmetric functions such as $f(\kappa)$ and $g(\kappa)$. In fact, the simplest answer seems to be:

$$j(\kappa) = \frac{(1 - \kappa + \kappa^2)^3}{\kappa^2(1 - \kappa)^2} \tag{17.23}$$

which obeys the symmetry idea as it can be re-expressed in the two forms:

$$j(\kappa) = \frac{[(1 - \kappa(1 - \kappa)]^3}{[\kappa(1 - \kappa)]^2} = \frac{(\kappa + 1/\kappa - 1)^3}{\kappa + 1/\kappa - 2} \tag{17.24}$$

Fortunately, we need to go no further in our quest to obey the six conditions required to recognize all six cross-ratio values κ, $(1 - \kappa)$, $1/\kappa$, $1/(1 - \kappa)$, $(\kappa - 1)/\kappa$, $\kappa/(\kappa - 1)$. The reason is that they are all deducible from each other by further applications of the initial negation and inversion rules. (The ultimate reason for this is that the operations to transform the function from one to another of the six forms form a group of order six, which is generated from the negation and inversion transforms.)

While this is a powerful result, it does not come without loss. The reason is that there is now a sixfold ambiguity inherent in the solution, so that once we have shown that the set of points satisfies the symmetric cross-ratio function, we still have to make tests to determine which of the six possibilities is the correct one. This is reflected by the complexity of the j-function, which contains a sixth degree polynomial and for every value of j there are six possible values of κ (Fig. 17.10).

The situation can be described by saying that the function $j(\kappa)$ is not "complete," in the sense that this function alone is insufficient to recognize the set of features unambiguously. To underline this, observe that the original cross-ratio is complete: once the value of κ is known, we can uniquely determine the position

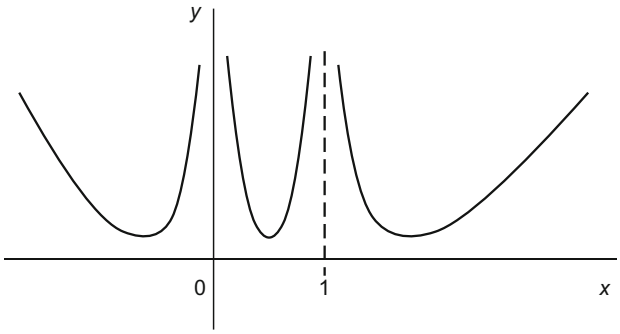

FIGURE 17.10

Symmetric cross-ratio function. This is the function defined by Eq. (17.23).

of one of the points from the other three points. This is obvious from the graph of κ as a function of x (where $x = x_{34}$ gives the position of the fourth point),[4] which is a hyperbola:

$$\kappa = \frac{x_{31}x_{24}}{x_{21}x_{34}} = \frac{x_{31}(x_{23}+x)}{x_{21}x} = \frac{x_{31}x_{23}}{x_{21}}\left(\frac{1}{x}+\frac{1}{x_{23}}\right) \tag{17.25}$$

17.7 VANISHING POINT DETECTION

In this section, we consider how vanishing points (VPs) can be detected. It is usual to carry this out in two stages: first, we locate all the straight lines in the image; next, we find which of the lines pass through common points. The latter being interpreted as VPs. Finding the lines using the Hough transform should be straightforward, although texture edges will sometimes prevent lines from being located accurately and consistently. Basically, locating the VPs requires a second Hough transform in which whole lines are accumulated in parameter space, leading to well-defined peaks (the VPs) where multiple lines overlap. In practice, the lines of votes will have to be extended to cover all possible vanishing point locations. This procedure is adequate when the VPs appear within the original image space, but it often happens that they will be outside the original image (Fig. 17.11) and may even be situated at infinity. This means that a normal image-like parameter space cannot be used successfully, even if it is extended beyond the original image space. Another problem is that for distant VPs, the peaks in parameter space will be

[4]In projective geometry, it is well known that there are three degrees of freedom on a line. The positions of three points on a line are not predictable from other views of the three points, without further information on the viewpoint.

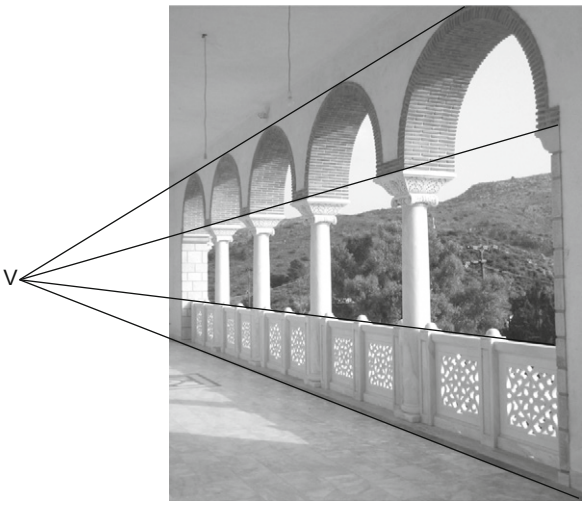

FIGURE 17.11

Position of the vanishing point. In this figure, parallel lines on the arches appear to converge to a vanishing point V outside the image. In general, vanishing points can lie at any distance and may even be situated at infinity.

spread out over a considerable distance, so detection sensitivity will be poor and accuracy of location will be low.

Fortunately, Magee and Aggarwal (1984) found an improved representation for locating VPs. They constructed a unit sphere G, called a Gaussian sphere, around the center of projection of the camera, and used G instead of an extended image plane as a parameter space. In this representation (Fig. 17.12), VPs appear at finite distances even in cases where they would otherwise appear to be at infinity. For this method to work, there has to be a one-to-one correspondence between points in the two representations, and this is clearly valid (note that the back half of the Gaussian sphere is not used). However, the Gaussian sphere representation is not without problems: in particular, many irrelevant votes will be cast from lines that are not parallel in real 3-D space (generally only a small subset of the lines in the image will pass through VPs). To solve this problem, *pairs* of lines are considered in turn, and their crossing points are only accumulated as votes if the lines of each pair are judged likely to originate from parallel lines in 3-D space (e.g., they should have compatible gradients in the image). This procedure drastically limits both the number of votes recorded in parameter space and the number of irrelevant peaks. Nevertheless, the overall cost is still substantial, being proportional to the number of pairs of lines. Thus, if there are N lines, the number of pairs is ${}^{N}C_2 = \frac{1}{2} N(N-1)$, so the result is $O(N^2)$.

The above procedure is important because it provides a highly reliable means for performing the search for VPs and largely discriminating against isolated lines

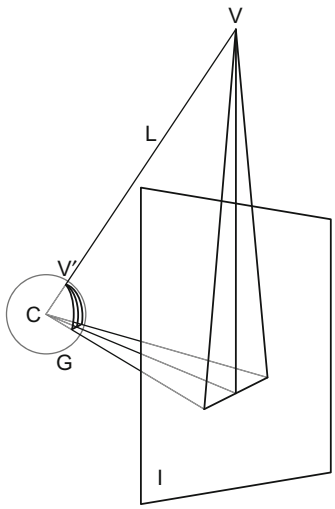

FIGURE 17.12

Detection of vanishing points using the Gaussian sphere. Parallel lines in space lead to converging lines in the image I. While the vanishing point V is here well above the image; it is easily located by projecting the lines onto the Gaussian sphere G. As discussed in the text, G is commonly used as a parameter space for accumulating vanishing point votes. C is the center of projection of the camera lens.

and image clutter. Note that for a moving robot or other system, the correspondences between the VPs seen in successive images will lead to considerably greater certainty in the interpretation of each image.

17.8 MORE ON VANISHING POINTS

One advantage of the cross-ratio is that it can turn up in many situations and on each occasion provide yet another neat result. A further example is when a road or pavement has flagstones whose boundaries are well demarcated and easily measurable. They can then be used to estimate the position of the VP on the ground plane. Imagine viewing the flagstones obliquely from above, with the camera or the eyes aligned horizontally (e.g., as for Fig. 23.12(a)). Then we have the geometry of Fig. 17.13 where the points O, H_1, H_2 lie on the ground plane, whereas O, V_1, V_2, V_3 are in the image plane.[5]

[5]Note that slightly oblique measurement of the flagstones, along a line that is not parallel to the sides of the flagstones, still permits the same cross-ratio value to be obtained, as the same angular factor applies to all distances along the line.

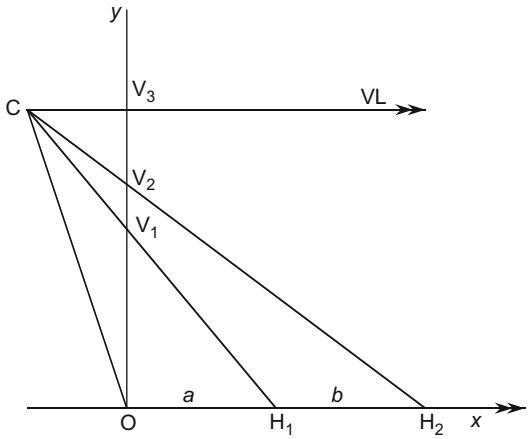

FIGURE 17.13

Geometry for finding the vanishing line from a known pair of spacings. C is the center of projection. VL is the vanishing line direction, which is parallel to the ground plane OH_1H_2. Although the camera plane $OV_1V_2V_3$ is drawn perpendicular to the ground plane, this is not necessary for successful operation of the algorithm (see text).

If we regard C as a center of projection, the cross-ratio formed from the points O, V_1, V_2, V_3 must have the same value as that formed from the points O, H_1, H_2, and infinity in the horizontal direction. Supposing that OH_1 and H_1H_2 have known lengths a and b, equating the cross-ratio values gives:[6]

$$\frac{y_1(y_3 - y_2)}{y_2(y_3 - y_1)} = \frac{x_1}{x_2} = \frac{a}{a+b} \tag{17.26}$$

This allows us to estimate y_3:

$$(a+b)(y_1y_3 - y_1y_2) = ay_2y_3 - ay_2y_1 \tag{17.27}$$

$$\therefore \quad y_3(ay_1 + by_1 - ay_2) = ay_1y_2 + by_1y_2 - ay_1y_2 \tag{17.28}$$

$$\therefore \quad y_3 = \frac{by_1y_2}{ay_1 + by_1 - ay_2} \tag{17.29}$$

If $a = b$ (as is likely to be the case for flagstones):

$$y_3 = \frac{y_1y_2}{2y_1 - y_2} \tag{17.30}$$

[6]Note that, in the case of Fig. 17.13, the y values are measured from O rather than from V_3.

Note that this proof does not actually assume that points V_1, V_2, V_3 are vertically above the origin, or that line OH_1H_2 is horizontal, just that these points lie along two coplanar straight lines and that C is in the same plane. Also, note that it is only the ratio of a to b, not their absolute values, that is relevant in this calculation.

Having found y_3, we have calculated the direction of the VP, whether or not the ground plane on which it lies is actually horizontal, and whether or not the camera axis is horizontal.

Finally, note that Eq. (17.30) can be rewritten in the simpler form:

$$\frac{1}{y_3} = \frac{2}{y_2} - \frac{1}{y_1} \tag{17.31}$$

The inverse factors give some intuition into the processes involved—not least considering the inverse relation $Z = Hf/y$ between distance along the ground plane and image distance from the vanishing line; and similarly, the inverse relation between depth and disparity in Eq. (15.4).

17.9 APPARENT CENTERS OF CIRCLES AND ELLIPSES

It is well known that circles and ellipses project into ellipses (or occasionally into circles). This statement is widely applicable and is valid for orthographic projection, scaled orthographic projection, weak perspective projection, and full perspective projection.

Another factor that can easily be overlooked is what happens to the center of the circle or ellipse under these transformations. It turns out that the ellipse (or circle) center does not project into the ellipse (or circle) center under full perspective projection: there will in general be a small offset (Fig. 17.14).[7]

If the position of the vanishing line of the plane can be identified in the image, the calculation of the offset for a circle is quite simple using the theory in Section 17.8, which applies as the center of a circle bisects its diameter (Fig. 17.15). First, let ε be the shift in the center, d the distance of the center of the ellipse from the vanishing line, and b the length of the semi-minor axis. Next, identify $b + \varepsilon$ with y_1, $2b$ with y_2, and $b + d$ with y_3. Finally, substitute for y_1, y_2, and y_3 in Eq. (17.30). We then obtain the result:

$$\varepsilon = \frac{b^2}{d} \tag{17.32}$$

[7]The fact that this happens may perhaps suggest that ellipses will be slightly distorted under projection. In fact, there is *no* such distortion, and the source of the shift in the center is merely that full perspective projection does not preserve length ratios—and in particular midpoints do not remain midpoints.

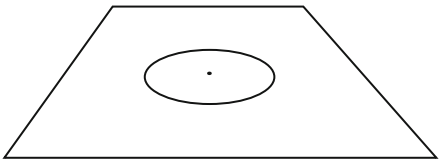

FIGURE 17.14

Projected position of a circle center under full perspective projection. Note that the projected center is not at the center of the ellipse in the image plane.

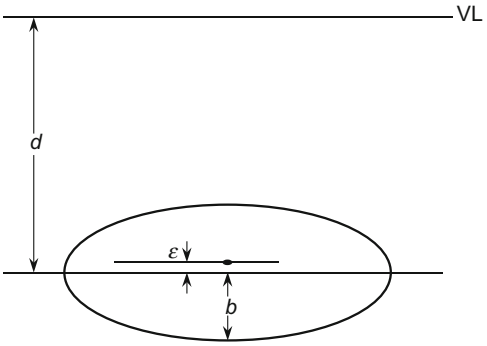

FIGURE 17.15

Geometry for calculating the offset of the circle center. The projected center of the circle is shown as the elongated dot, and the center of the ellipse in the image plane is a distance *d* below the vanishing line VL.

Note that, unlike the situation in Section 17.8, we are here assuming that y_3 is known and we are using its value to calculate y_1 and hence ε.

If the vanishing line is not known, but the orientation of the plane on which the circle lies is known, and also the orientation of the image plane, then the vanishing line can be deduced, and the calculation can again proceed as above. However, this approach assumes that the camera has been calibrated (see Chapter 18).

The problem of center determination when an ellipse is projected into an ellipse is a little harder to solve: not only is the longitudinal position of the center unknown, but so is the lateral position. Nevertheless, the same basic projective ideas apply. Specifically, let us consider a pair of parallel tangents to the ellipse, which in the image become a pair of lines λ_1, λ_2 meeting on the vanishing line (Fig. 17.16). As the chord joining the contact points of the tangents passes through the center of the original ellipse, and as this property is projectively invariant, the projected center must lie on the chord joining the contact points of

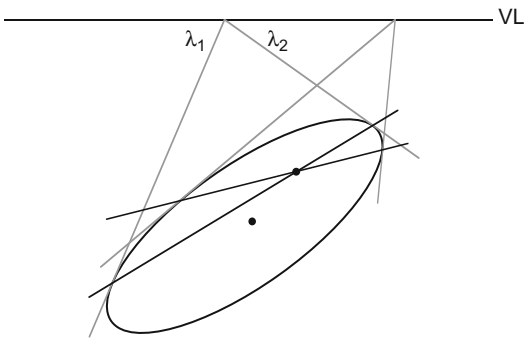

FIGURE 17.16

Construction for calculating the offsets for a projected ellipse. The two lines λ_1, λ_2 from a point on the vanishing line VL touch the ellipse, and the joins of the points of contact for all such line pairs pass through the projected center. (The figure shows just two chords of contact.)

the line pair λ_1, λ_2. As the same applies for all pairs of parallel tangents to the original ellipse, we can straightforwardly locate the projected center of the ellipse in the image plane.[8] For an alternative, numerical analysis of the situation, see Zhang and Wei (2003).

Both the circle result and the ellipse result are important in cases of inspection of mechanical parts, where accurate results of center positions have to be made irrespective of any perspective distortions that may arise. Indeed, circles can also be used for camera calibration purposes, and again high accuracy is required (Heikkilä, 2000).

17.10 THE ROUTE TO FACE RECOGNITION

In a book of this type, it is important to consider the problem of face recognition, which has long been a target for many practitioners of machine vision. In fact, some would say that by now this is a solved problem, but criminologists repeatedly confirm that there is still some distance to go before reliable identification of people can be attained in practical circumstances. Not least, there are problems of

[8]Students who are familiar with projective geometry will be able to relate this to the "pole–polar" construction for a conic: in this case the polar line is the vanishing line and its pole is the projected center. In general the pole is not at the center of an ellipse, and will not be so unless the polar is at infinite distance. Indeed, from this point of view, Eq. (17.31) can be understood in terms of "harmonic ranges," y_2 being the "harmonic mean" of y_1 and y_3.

hairstyles, hats, glasses, beards, degree of facial stubble, wildly variable facial expressions, and of course variations in lighting and shadow—and this does not even touch on the problem of deliberate disguise. Added to this, one must never forget that the face is not flat, but part of a solid, albeit malleable object—the head—which can appear in a variety of orientations and positions in space.

These remarks make it clear that full analysis and solution of the face recognition problem would require a whole book on the topic, and several such books have appeared over time (e.g., Gong et al., 2000). However, the topic of facial recognition is itself dated: not only are workers interested in face recognition, but also there is pressure to measure facial expression for reasons as diverse as determining whether a person is telling the truth and finding how to mimic real people as accurately as possible in films (over the next few years, it is possible to anticipate that a good proportion of films will contain no human actors as this has the potential for making them quicker and cheaper to produce). Clearly, medical diagnosis or facial reconstruction can also benefit from facial measurement algorithms, while person verification may well be at least as important as face recognition, so long as it can be done quickly and with minimum error. In this latter respect, recent efforts have moved in the direction of identifying people highly accurately from their iris patterns (e.g., Daugman, 1993, 2003), and even more accurately from their retinal blood vessels[9] (using the methods of retinal angiography). While retinal methods would be rather expensive to implement, e.g. on all-weather ATM machines, the iris method need not be, and much progress has been made in this direction.

These considerations show that a narrow view of face recognition would be rather inappropriate. For this reason we concentrate here on one or two important aspects. Among these are the task of analyzing the face for key features that can then at least provide a proper framework for further work on facial recognition, facial expression, facial verification, and so on, and the task of locating the iris, which will be used both for detailed verification and as an important starting point for facial analysis. In fact, location of the iris can be dealt with reasonably straightforwardly using the Hough transform approach. This has already been covered in Section 12.7.

Other facial features, such as the corners of the eye and mouth, can be found by tracking, snake algorithms, or simply by corner detection. Similarly, the upper and lower contours of the ear and of the nose can be ascertained, thus yielding fixed points that can be used for a multitude of purposes ranging from person identification to recognition of facial expressions. At this point, it is useful to reconsider the fact that the face is part of the head, and that this is a 3-D object.

[9]Here some of the main deciding factors are commercial rather than academic, although an important message is that the technical difficulty is only viable where there is a need for the highest security, but in that case "the false acceptance rate for a correctly installed retina scan system falls below 0.0001 percent" (http://ru.computers.toshiba-europe.com; website accessed May 19, 2004).

17.10.1 The Face as Part of a 3-D Object

To start an analysis of the head and face, note that we can define a plane Π containing the outer corners of the eyes and the outer corners of the mouth, if we assume that some odd facial expression has not been adopted. To a very good approximation, it can also be assumed that the inner corners of the eyes will be in the same plane (Fig. 17.17(a–c)). The next step is to estimate the position of the vanishing point V for the three horizontal lines λ_1, λ_2, λ_3 joining these three pairs of features (Fig. 17.17(d)). Once this has been done, it is possible to use the relevant cross-ratio invariants to determine the points that in 3-D lie midway between the two features of each pair. This will give the symmetry line λ_s of the face. It will also be possible to determine the horizontal orientation θ of the facial plane Π, i.e., the angle through which it has been rotated, about a vertical axis, from a full frontal view. The geometry for these calculations is shown in Fig. 17.17(e). Finally, it will be possible to convert the inter-feature distances along λ_1, λ_2, λ_3 to the corresponding full frontal values, taking proper account of perspective, but not yet taking account of the vertical orientation φ of Π, which is still unknown.[10]

In fact, there is insufficient information to estimate the vertical orientation φ, without making further assumptions. Ultimately, this is because the face has no horizontal axis of symmetry. If we can assume that φ is zero (i.e., the head is held neither up nor down, and the camera is on the same level), then we can gain some information on the relative vertical distances of the face, the raw measurements for these being obtained from the intercepts of λ_1, λ_2, λ_3 with the symmetry line λ_s. Alternatively, we can assume average values for the inter-feature distances and deduce the vertical orientation of the face. A further alternative is to make other estimates based on the chin, nose, ears, or hairline, but as these are not guaranteed to be in the facial plane Π, the whole assessment of facial pose may not then be accurate and invariant to perspective effects.

Overall, we are moving toward measurements either of facial pose or of facial inter-feature measurements, with the possibility of obtaining some information on both, even when perspective distortions have to be allowed for (Kamel et al., 1994; Wang et al., 2003). Of course, the analysis will be significantly simpler in the absence of perspective distortions, when the face is viewed from a distance or when a full frontal view is guaranteed. Indeed, the bulk of the work on facial recognition and pose estimation to date has been done in the context of weak perspective, making the analysis altogether simpler. Even then, the possibility of wide varieties of facial expression brings in a great deal of complexity. It should also be noted that the face is not merely a rubber mask (or deformable template) that can be distorted "tidily". The capability for opening and closing the mouth and eyes creates additional nonlinear effects that are not modeled merely by variable stretching of rubber masks.

[10]Note that the theory underlying these procedures is closely related to that of Section 17.8. See also Fig. 17.13.

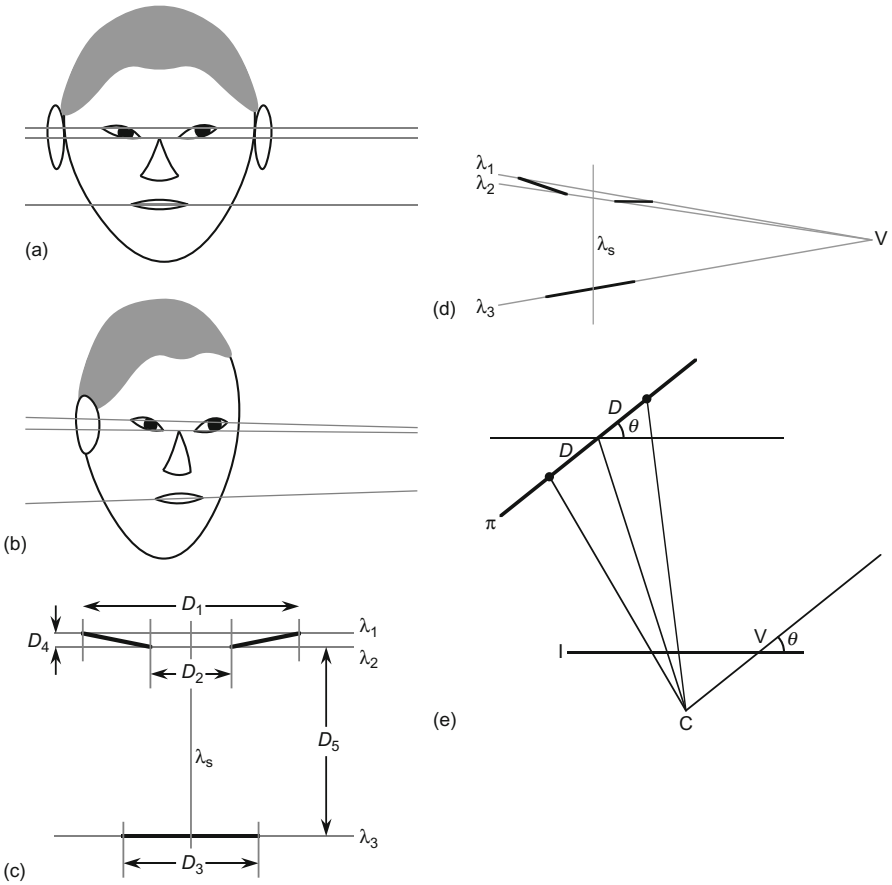

FIGURE 17.17

3-D analysis of facial parameters. (a) Front view of face. (b) Oblique view of face, showing perspective lines for corners of the eye and mouth. (c) Labeling of eye and mouth features and definition of five inter-feature distance parameters. (d) Position of vanishing point under an oblique view. (e) Positions of one pair of features and their midpoint on the facial plane Π. Note how the midpoint is no longer the midpoint in the image plane I when viewed under perspective projection. Note also how the vanishing point V gives the horizontal orientation of Π.

17.11 PERSPECTIVE EFFECTS IN ART AND PHOTOGRAPHY*

An artist is painting a picture somewhere on the countryside. Every now and again he looks up from his easel and surveys the scene, then he turns back to his picture and adds a few more touches. He has chosen his location carefully, and has set his easel at the right angle for best effect. We will suppose that he is not aiming to be impressionist, but wishes to present the scene as he sees it. Although his painting is in 2-D, he is able to present all the information needed for others to perceive the scene in 3-D. However, there is a problem. The picture needs to be viewed from precisely the right angle and distance that must correspond to the artist's original viewpoint. Of course, the artist had to rotate his head and mentally rotate the scene between the moments that he viewed it and painted it (he had to do this because the canvas he was painting on was opaque: other artists such as Canaletto have used camera obscura methods to overcome this difficulty). However, we can overcome the problem by temporarily assuming that the canvas is transparent, which significantly simplifies the geometry (Fig. 17.18).

Interestingly, from his viewpoint, the artist could have painted a whole range of pictures of the scene, with his easel set at different angles (Fig. 17.18). All these pictures would be very closely related to each other and in fact would be related by homographies. But each of them would have exactly one proper viewing position and orientation, and when each was viewed from its proper viewing position, exactly the same 3-D regenerative effect would be perceived by the viewer. Thus,

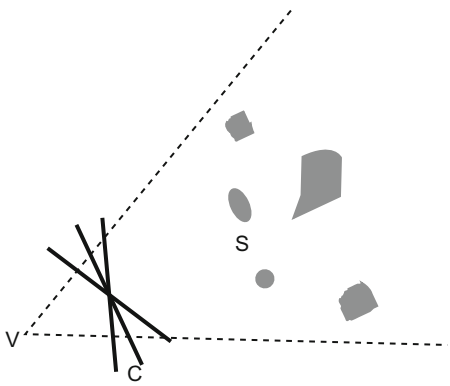

FIGURE 17.18

Effective viewpoint of artist painting a picture. The artist views the scene S from viewpoint V, and paints what he sees on the canvas at C. The picture painted at C could be one of many depending on the orientation of the canvas.

*Parts of this Section were inspired by a TV lecture dated 28 December 1978 by Christopher Zeeman entitled "Mathematics into pictures".

the fact that a homography exists between the various views does not change the constraint that each version of a picture is best viewed from a single location.

However, there is a circumstance when this is no longer true. That is, when the scene contains a flat (2-D) surface F that is then to be presented in 2-D. Immediately, we have a homography between the original scene and the canvas, and we also have homographies with all the possible rotated versions of the canvas. But what of the viewing positions? To understand the situation properly, we need to think of the possible viewing points relative to the frame of reference of the canvas C, which we must now regard as fixed, e.g., on a gallery wall W. We can see that as the original canvas is rotated, so the ideal viewing point relative to its location on the gallery wall will rotate, albeit remaining at the distance corresponding to the distance of the artist's eye from the canvas. In fact, as one walks (along a circle) around the painting in the gallery, all the possible pictures that the artist might have painted from his original position unfold before us (Fig. 17.19). They all embody valid perspective distortions and thus all would appear entirely natural. Note that a circular path is appropriate because it corresponds to the (constant) overall angle of the artist's view (angles in the same segment of a circle are equal).

But what of the case not of the flat wall of a house but of a face? In fact, substantial parts of the face approximate to a flat 2-D surface, e.g., the forehead, eyes, cheeks, mouth, and chin. Considering them alone, a considerable range of viewing points would be acceptable. Then there is the human propensity for focussing on the eyes and largely ignoring the rest of the face. If this is done, acceptable views will be obtained by viewing the painting from many directions. Indeed when focussing on the eyes and ignoring the rest, it seems entirely

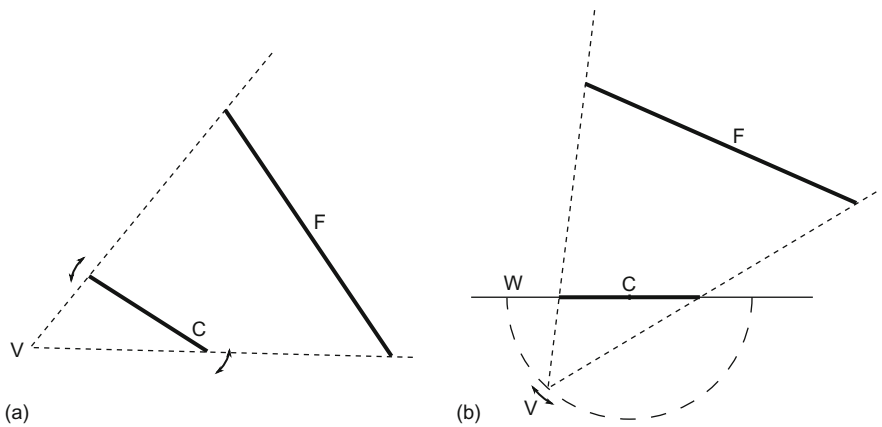

FIGURE 17.19

Viewing a painted picture. Process of (a) painting and (b) viewing a picture. As the orientation of the canvas changes in (a), so the proper viewing point V in (b) moves along the path of a circle. For a flat object F, the circular path sweeps out all possible pictures the artist might have painted, and all will be related by homographies (see text).

understandable why people would report after visiting a stately home and seeing a painting of the 17th Earl, that his eyes "followed them around the room."

A further factor is involved in this analysis—the orientation of the face when the picture was painted. If the face was originally at an angle α, the eyes would appear a factor $\cos\alpha$ closer together than in the head-on painting. However, if the canvas were rotated through an angle β, the eyes would be enlarged by a factor $\sec\beta$. Hence, there would be an overall magnification $\cos\alpha/\cos\beta$. Cancellation would occur when $\beta = \alpha$, which corresponds to the canvas being parallel to the face. However, cancellation would also occur when $\beta = -\alpha$, and this corresponds to the canvas and the face being rotated through equal and opposite angles α relative to the final viewing direction (Fig. 17.20). Next, suppose that a certain amount of enlargement or diminution of the apparent distance between the eyes is acceptable (note especially that if one doesn't know the 17th Earl in the painting, some enlargement would be acceptable). The result is that the range of acceptable orientations of the final viewing direction will be increased, with the β distortion tending to cancel the α distortion, the largest distortions for given α occurring at high or low $|\beta|$ (Fig. 17.21). Equalizing these extreme distortions by making $|\alpha| > 0$ would give the maximum permissible range of acceptable orientations (e.g., $\alpha = 20°$, $|\beta| = 0 - 40°$ with $|\beta| = \alpha$ in the middle of the range).

In photography there is also a correct viewing position, but when examining family photographs, there are no exceptional situations where people would look only at the eyes: people would want to look at facial expressions, hairstyles, and so on (they would also be quite sensitive to whether everyone's eyes were open). Unfortunately, photographs of groups often appear distorted around the outside, a factor that could sometimes be partly due to pincushion or barrel distortion (these are lens aberrations—see Chapter 18). However, this effect could also be due to an incorrect viewing position: the camera doesn't lie, but only shows the true

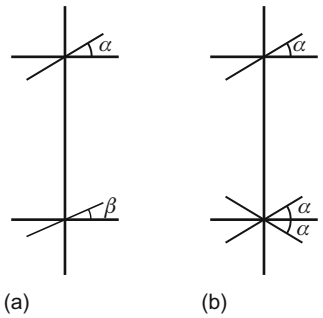

FIGURE 17.20

Effect of rotations relative to the viewing direction. In (a), α represents the original orientation of the face, and β represents the orientation of the canvas. Part (b) shows the situation both when $\beta = \alpha$ and when $\beta = -\alpha$. In both instances, the two orientation effects cancel and the eyes appear to have their original separation.

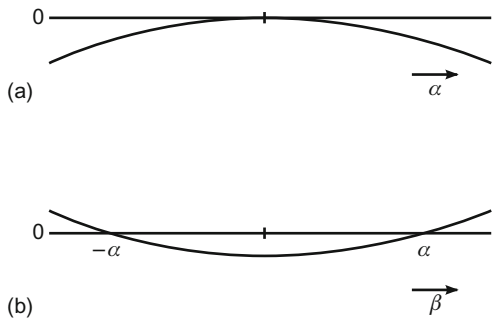

FIGURE 17.21

Effect of varying the viewing orientations. Part (a) shows how the separation of the eyes becomes reduced as $|\alpha|$ increases from zero. Part (b) shows how the change in separation of the eyes for a fixed value of α is first canceled out by increasing $|\beta|$ from zero and then increased so that the change in separation becomes positive. The average magnitude of the change in separation can clearly be lower in (b) than in (a).

geometry according to the perspective it had when the button was pressed. It is a fact that photographs are often viewed at arms length—a distance far greater than the correct viewing distance (Fig. 17.22). If photographs were meant to be viewed at that distance, the photograph should be taken from much further away, to be sure that the camera doesn't lie inadvertently. Of course, there is a complication that people will normally look at the camera, so taking the photograph from a different distance will affect the material content of the scene itself.

Until quite recently, photographs were best taken close up, because of the limited resolution of the film. Nowadays, digital cameras have such phenomenal resolution that there is some advantage from taking pictures further away, or even from a considerable distance, with the aid of a zoom lens. (The latter confers the added advantage that real-life shots can be taken without the subject being embarrassed or even aware they are being taken.) But there is a quite different advantage to be gained from taking photographs much further away: although the correct viewing distance could then be rather larger than ideal from the perspective point of view, all people at all locations in the photograph can be viewed individually without perspective distortions creeping in. Note again that it is common practice for photographs to be handed around, and for each person in them to be scrutinized *individually*—so the overall global composition might well be less important than the individual people who are portrayed (this is all the more true when one knows the people in the photograph, which is much more likely than with a painting of the 17th Earl). Optimization not only of each locality in a photograph but also of the global view is plainly impossible, but taking the photograph from a distance gives a very good compromise (Fig. 17.22(b)). However, taking it from infinity would lead to zero foreshortening of the faces and thereby make them appear flatter. Here a lot depends on whether the lighting provides other cues that can give a good impression of depth.

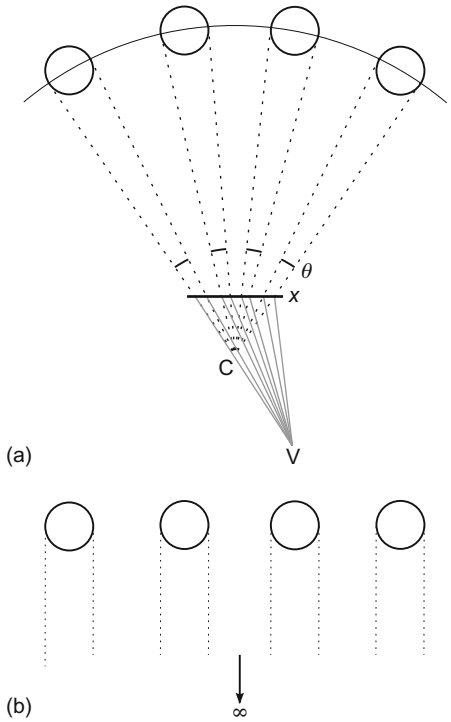

FIGURE 17.22

Process of taking and viewing a photograph. (a) The geometry for taking a photograph of a group of people all facing the camera at C. Also shown is the effective viewing location V of the photograph. (Clearly, the photograph would be enlarged before viewing, in which case the part of the illustration just above V would be scaled, but not otherwise changed.) (b) A potentially more ideal way of taking the photograph, from a large distance. By following (b), people examining the photograph would be viewing from an ideal viewing point; in addition, all people shown in the picture could be examined individually without distortion.

Another possibility that digital photography offers is that of automatically stitching together several frames to create a wide scene or panorama. Here best results are obtained if the camera is put into stitching mode, so that it can make exposures constant over a sequence of frames or at least record what they were. Then it could be expected that the edges of various frames would match up without any sudden changes between them. To achieve a proper match, the frames obviously have to overlap, and then special software can be used to find the best set of lines for trimming and stitching. This involves moving the trim lines in such a way—generally in plain background regions—that the breaks will be

imperceptible. Of course, some smoothing along the trimmed edges will often help, as long as this process doesn't encroach on regions of totally different intensity. Unfortunately, stitching cannot easily cope with situations containing moving objects, and in this context the author had an interesting early experience of the final country scene containing too many sheep.

The exact algorithms to be used for image stitching are quite intricate. This is because a time-consuming search is required to identify the best trimming lines. The criteria to be used are important, but these obviously involve minimization of intensity and color change along the borders—and there is also advantage in minimizing the intensity and color *gradients* along the borders. Then there are rules for smoothing along the borders, after stitching has been carried out. Here the simplest rule is not to do any smoothing at all, but this rule can be relaxed if the changes in intensity and intensity gradient have successfully been minimized.

There is one major problem with stitching—that scenes containing straight edges that go at angles across the whole scene are almost impossible to deal with (Fig. 17.23). This is because each (flat) frame will have been taken from a different direction in order to obtain a wider overall view. Thus, a straight line, such as a path, will appear straight in each frame, but the orientation will normally have to change at the join (Fig. 17.23): that this is not mere theory is demonstrated clearly by the examples in Fig. 17.24. The only way to overcome the problem is to present the scene on a sphere or cylinder in order to prevent kinks from appearing at the joins. But then the original straight line will become a curved line, especially when the final picture is presented as a flat scene. Here again the specter of the single viewpoint of each original picture is upon us. The best way of handling it seems to be to present it as a picture that is apparently taken from infinity (as for the earlier example of photographing a group of people), so that any straight line will appear straight, at least at every local position, even if a ruler placed along it will show that it is not straight globally. In fact, this is what the special rotating line-scan cameras were able to achieve in the 1960s and earlier, when they were used to take school photographs—and the time taken to gather enough light was often sufficient for at least one small boy to run from one end to the other and be photographed twice!

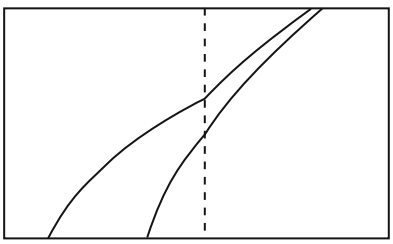

FIGURE 17.23

Effect of stitching two pictures depicting parts of a path. Apparently correct stitching will actually lead to a kink at the join.

FIGURE 17.24

Practical instances of stitching. Part (a) shows that the surmise of Fig. 17.23 is correct. Part (b) shows the result of using an (effectively) overenthusiastic stitching package that manages to avoid a kink, but ends up with perspective nonsense. If the lower end of the road boundary had been visible, the software might have avoided the latter problem but would then have introduced a kink.

17.12 CONCLUDING REMARKS

This chapter has aimed to give some insight into the important subject of invariants and its application in image recognition. The subject takes off when ratios of ratios of distances are considered, and this idea leads in a natural way to the cross-ratio invariant. While its immediate manifestation lies in its application to recognition of the spacings of points on a line, it generalizes immediately to angular spacings for pencils of lines and also to angular separations of concurrent planes. A further extension of the idea is the development of invariants that can describe sets of noncollinear points, and two cross-ratios suffice to characterize a set of five noncollinear points on a plane. The cross-ratio can also be applied to conics. Indeed, Chasles' theorem describes a conic as the locus of points that maintains a pencil of constant cross-ratio with a given set of four points. However, this theorem does not permit one type of conic curve to be distinguished from another.

Many other theorems and types of invariant exist, but space prevents more than a mention being made of them. As an extension to the line and conic

examples given in this chapter, invariants have been produced that cover a conic and two coplanar nontangent lines, a conic and two coplanar points, and two coplanar conics. Of particular value is the group approach to the design of invariants (Mundy and Zisserman, 1992a). However, certain mathematically viable invariants, such as those that describe local shape parameters on curves, prove to be too unstable to use in their full generality because of image noise. Nevertheless, semi-differential invariants have been shown (Section 17.5) to be capable of fulfilling essentially the same function.

Next, there is the warning of Åström (1995) that perspective transformations can produce such incredible changes in shape that a duck silhouette can be projected arbitrarily closely into something that looks like a rabbit or a circle, hence upsetting invariant-based recognition.[11] While such reports seem absent from the previous literature, Åström's work indicates that care must be taken to regard recognition via invariants as hypothesis formation that is capable of leading to false alarms.

Overall, the value of invariants lies in making computationally efficient checks of whether points or other features might belong to specific objects. In addition, they achieve this without the necessity[12] for camera calibration or knowing the viewpoint of the camera (although there is an implicit assumption that the camera is Euclidean). While invariants have been known within the vision community for well over 20 years, it is only during the last about 15 years that they have been systematically developed and applied for machine vision. Such is their power that they will undoubtedly assume a stronger and more central role in the future. Nowhere is this power better indicated than by the application to vanishing point detection and face recognition described in Sections 17.7–17.10. Note also the perspective projection problems that led not only to the need for invariants but also to the need for further insight into the problems of viewing and stitching 2-D pictures (Section 17.11).

The problems of perspective projection are ubiquitous in 3-D vision, appearing even in simple situations such as viewing 2-D pictures and stitching digital photographs. However, vital interpretive information is provided by projective invariants that slice right through such complexities and are able to help, e.g., with vanishing point detection and face recognition.

[11] It could of course be argued that all recognition methods will be subject to the effects of perspective transformations. However, invariant-based recognition will not flinch from invoking highly extreme transformations that appear to grossly distort the objects in question, whereas more conventional methods are likely to be designed to cope with a reasonable range of expected shape distortions.

[12] Here we assume that the aim is location of specific objects in the image. If the objects are then to be located in the world coordinates, some form of camera calibration will again be needed. However, there are many applications, such as inspection, surveillance, and identification (e.g., of faces or signatures) where location of objects in the *image* can be entirely adequate.

17.13 BIBLIOGRAPHICAL AND HISTORICAL NOTES

The mathematical subject of invariance is very old (cf. the work of Chasles, 1855), but it has only relatively recently been developed systematically for machine vision. Notable in this context is the work of Rothwell, Zisserman, and their coworkers, as reported by Forsyth et al. (1991), Mundy and Zisserman (1992a,b), Rothwell et al. (1992a,b), and Zisserman et al. (1990). In particular, the paper by Forsyth et al. (1991) shows the range of available invariant techniques and discusses the problems of stability which arise in certain cases. The appendix (Mundy and Zisserman, 1992b) on projective geometry for machine vision, which appears in Mundy and Zisserman (1992a), is especially valuable, and provides the background needed for understanding the other papers in the volume. On the whole, the latter volume has a theoretical flavor that demonstrates what ought to be possible using invariants, although comparisons between invariants and other approaches to recognition are perhaps lacking. Thus, it is only by examining whether workers choose to use invariants in real applications that the full story will emerge. In this respect, the paper by Kamel et al. (1994) on face recognition is of great interest, as it shows how invariants helped to achieve more than had been achieved earlier after many attempts using other approaches—specifically in correcting for perspective distortions during face recognition.

Other more recent work appears in a special issue of *Image and Vision Computing* (Mohr and Wu, 1998). In particular, the paper by Van Gool et al. (1998) shows how shadows can be allowed for in aerial images, and the paper by Boufama et al. (1998) shows how invariants can help with object positioning. Startchik et al. (1998) provides a useful demonstration of the semi-differential invariant methods covered in Section 17.5. Maybank (1996) deals with the problem of accuracy with invariants, making the point that this can be serious even for cross-ratios (which contain only four parameters). Another early work, by a totally different set of workers, is Barrett et al. (1991) and contains a number of useful derivations, together with a practical example of aircraft recognition, complete with accuracy assessments.

Rothwell's (1995) book covers the early work on invariance in a thoughtful manner, and the later 3-D books by Hartley and Zisserman (2000) and Faugeras and Luong (2001) integrate the ideas into their structure, but are not always easy to understand by students starting out in the subject. Semple and Kneebone (1952) is a standard work on projective geometry, which is still widely used in its later reprints.

Vanishing point determination has often been considered both in relation to egomotion for mobile robots (Lebègue and Aggarwal, 1993; Shuster et al., 1993) and in general with regard to the vision methodology (Magee and Aggarwal, 1984; Shufelt, 1999; Almansa et al., 2003), which is prone to suffer from inaccuracy when real off-camera data are used in any context. The seminal paper that gave rise to the crucial Gaussian sphere technique was that by Barnard (1983). In an interesting twist, Clark and Mirmehdi (2002, 2003) use VPs to recover text

that has been distorted by perspective. The approach permits them to recover paragraph formats; in addition to line spacings, various forms of text justification can be recognized and managed.

17.13.1 More Recent Developments

More recently, Shioyama and Uddin (2004) have used cross-ratio invariants for the reliable location of pedestrian crossings by analyzing multiple crossing points of transverse straight lines with the alternating patterns on the road. Kelly et al. (2005) have used homographies between stereo views to locate shadows and low-lying objects: to achieve this they used the direct linear transformation (see Hartley and Zisserman, 2003) to identify homographies for sets of four or more points. Once homographies are found, they are used to eliminate the corresponding objects from consideration, thereby avoiding costly computation of 3-D depth values from the stereo views for those objects. Rajashekhar et al. (2007) use cross-ratio values to identify man-made structures in images to aid image retrieval. Hough transforms are used to find line structures in images, then feature points on the lines are found and sets of cross-ratio values are computed and presented in the form of histograms (in each case, all six possible cross-ratio values are included in the histograms). It is found that values in the range 0−5 are most suitable for identifying man-made structures, in that the histograms are suitably densely concentrated. Structures such as buildings are well recognized from the histograms, as long as they are quantized with upward of 200 bins. Li and Tan (2010) use a similar approach, but their cross-ratio values occur in continuous streams as outlines of characters or symbols are tracked. The resulting "cross-ratio spectra" allow characters to be recognized even with severe perspective distortions.

In the area of face recognition, An et al. (2010) describe a new illumination normalization model that is able to cope with varied lighting conditions. It works by decomposing the face into a high-frequency part and a low-frequency part: the main innovation is to divide the original intensity pattern by a smoothed version of the low-frequency part (although several other equalization and normalization processes are carried out as well). Hansen and Ji (2010) survey models for eye detection and gaze estimation and summarize the developments that are still needed in this area. Fang et al. (2010) describe a new method of multiscale image stitching. The paper discusses the problems of obtaining global and local alignment. A number of strategies are needed to overcome the various problems, and an iterative processing pipeline is required to integrate the different strategies.

17.14 PROBLEMS

1. Show that the six operations required to transform the cross-ratio κ into the six different values for four points on a line form a group G of order six (see Sections 17.2 and 17.6). Show that G is a noncyclic group, and has two

subgroups of order 2 and 3, respectively. *Hint:* Show that all possible combined operations fall within the same set of six, and also that this set contains the identity operation and the inverses of all the elements of the set.

2. Show that a conic and two points can be used to define an invariant cross-ratio.

3. Show that two conics can be used to define an invariant cross-ratio (a) if they intersect in four points, (b) if they intersect in two points, (c) and if they do not intersect at all, so long as they have common tangents.[13]

4. a. Perform a geometric calculation based on the sine rule, which shows that the angles α, β, γ are related to the distances a, b, c in Fig. 17.25 by the equation:

$$\frac{a}{\sin\alpha} \times \frac{c}{\sin\gamma} = \frac{a+b}{\sin(\alpha+\beta)} \times \frac{b+c}{\sin(\beta+\gamma)}$$

b. Show that this leads to a relation between the cross-ratios for various distances on the line and for the sines of various angles. Hence, show that this also leads to the constancy of the cross-ratios on any two lines crossing the pencil of four lines passing through O.

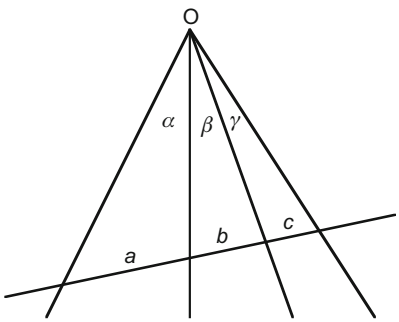

FIGURE 17.25

Geometry for cross-ratio calculation.

5. a. Explain the value of using *invariants* in relation to pattern recognition systems. Illustrate your answer by considering the value of thinning algorithms in optical character recognition.

b. The *cross-ratio* of four points (P_1, P_2, P_3, P_4) on a line is defined as the ratio:

$$C(P_1, P_2, P_3, P_4) = \frac{(x_3 - x_1)(x_2 - x_4)}{(x_2 - x_1)(x_3 - x_4)}$$

[13]The case of nonintersecting conics with no common tangents requires complex algebra. See, e.g., Rothwell (1995). The possibility of ambiguity and incompleteness is also discussed in Rothwell (1995).

Explain why this is a useful type of invariant for objects viewed under full perspective projection. Show that labeling the points in reverse order will not change the value of the cross-ratio.

c. Give arguments why the cross-ratio concept should also be valid for weak perspective projection. Work out a simpler invariant that is valid for straight lines viewed under weak perspective projection.

d. A flat lino-cutter blade has two parallel sides of different lengths. It is viewed under weak perspective projection. Discuss whether it can be identified from any orientation in 3-D by measuring the lengths of its sides.

6. a. Flagstones are viewed on a pavement, providing a large number of coplanar feature points. Show that a correspondence can be made between five coplanar feature points in two images—however, the camera has been moved between the shots—by checking the values of two cross-ratios.

b. It is required to compose a panorama of a scene by taking a number of photographs and "stitching" them together after making appropriate image transformations. To achieve this, it is necessary to make correspondences between the images. Show that the two cross-ratios type of planar invariant can be used for this purpose, even if the chosen scene features do not lie on a common plane. Determine under what conditions this is possible.

7. Redraw Fig. 17.16 using vanishing points aligned along the observed ellipse axes. Show that the problem of finding the transformed center location now reduces to two 1-D cases and that Eq. (17.32) can be used to obtain the transformed center coordinates.

8. A robot is walking along a path paved with rectangular flagstones. It is able to rotate its camera head so that one set of flagstone lines appears parallel while the other set converges toward a vanishing point. Show that the robot can calculate the position of the vanishing point in two ways: (1) by measuring the varying widths of individual flagstones or (2) by measuring the lengths of adjacent flagstones and proceeding according to Eq. (17.30). In the first case, obtain a formula that could be used to determine the position of the vanishing point. Which is the more general approach? Which would be applicable if the flagstones appeared in a flower garden in random locations and orientations?