

16

Tackling the Perspective n -point Problem

It is possible to recognize 3-D objects from very few point features, even when they are seen in a single view. In fact, the pose of the 3-D object can also be ascertained from a single view. However, ambiguities of interpretation do arise and this chapter discusses the disambiguation problem.

Look out for:

- the distinction between weak and full perspective projection.
- how the “perspective inversion” type of ambiguity arises under weak perspective projection.
- how more serious ambiguities arise under full perspective projection.
- how full perspective projection has the capability to provide more interpretative information than weak perspective projection.
- how coplanarity can impose quite strong constraints on 3-D data, which can be sometimes helpful and at other times an impediment.
- how symmetry can help with 3-D image interpretation.

Note that while this chapter considers only one aspect of 3-D vision, it raises very important issues that are relevant right through the subject of 3-D object recognition.

16.1 INTRODUCTION

This chapter follows on from the previous introductory chapter, and tackles a problem of central importance in the analysis of images from 3-D scenes. It has been kept separate and fairly short so as to focus carefully on relevant factors in the analysis. First, we look at the phenomenon of perspective inversion, which has already been alluded to several times in Chapter 15. Then we refine our ideas on perspective, and proceed to consider the determination of object pose from

salient features that are located in the images. It will be useful to consider how many salient features are required for unambiguous determination of pose.

16.2 THE PHENOMENON OF PERSPECTIVE INVERSION

In this section, we study first the phenomenon of *perspective inversion*. This is actually a rather well-known effect that appears in the following “Necker cube” illusion. Consider a wire cube made from 12 pieces of wire welded together at the corners. Looking at it from approximately the direction of one corner, it is difficult to tell which way round the cube is, i.e., which of the opposite corners of the cube is the nearer (Fig. 16.1). Indeed, on looking at the cube for a time, one gradually comes to feel one knows which way round it is, but then it suddenly appears to reverse itself; then that perception remains for some time, until it too reverses itself.¹ This illusion reflects the fact that the brain is making various hypotheses about the scene, and even making decisions based on incomplete evidence about the situation (Gregory, 1971, 1972).

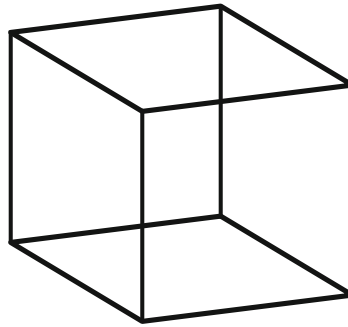
The wire cube illusion could perhaps be regarded as somewhat artificial. But consider instead an aeroplane (Fig. 16.2(a)) that is seen in the distance (Fig. 16.2(b)) against a bright sky. The silhouetting of the object means that its surface details are not visible. In that case interpretation requires that a hypothesis be made about the scene, and it is possible to make the wrong one. Clearly (Fig. 16.2(c)), the aeroplane could be at an angle α (as for P), although it could equally well be at an angle $-\alpha$ (as for Q). The two hypotheses about the orientation of the object are related by the fact that the one can be obtained from the other by reflection in a plane R normal to the viewing direction D.

Strictly, there is only an ambiguity in this case if the object is viewed under orthographic or scaled orthographic projection.² However, in the distance, perspective projection approximates to scaled orthographic projection, and it is often difficult to detect the difference.³ If the aeroplane in Fig. 16.2 were quite near, it would be obvious that one part of the silhouette was nearer, as the perspective would distort it in a particular way. In general, perspective projection will break down symmetries, so searching for symmetries that are known to be present in the object should reveal which way around it is: however, if the object is in the distance, as in Fig. 16.2(b), it will be virtually impossible to see the breakdown.

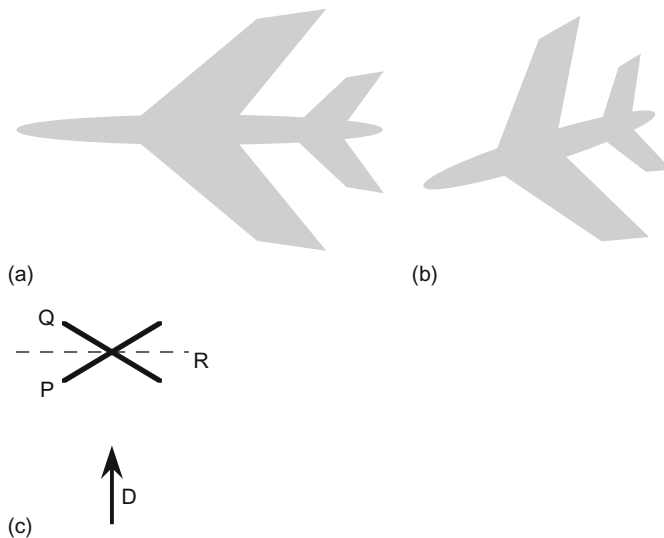
¹In psychology, this shifting of attention is known as *perceptual reversal*, which is unfortunately rather similar to the term *perspective inversion*, but is actually a much more general effect that leads to a host of other types of optical illusion—see Gregory (1971) and the many illustrations produced by M.C. Escher.

²*Scaled orthographic projection* is orthographic projection with the final image scaled in size by a constant factor.

³In this case, the object is said to be viewed under *weak perspective projection*. For weak perspective, the depth ΔZ within the object has to be much less than its depth Z in the scene. On the other hand, the perspective scaling factor can be different for each object and will depend on its depth in the scene: so the perspective can validly be locally weak and globally normal.

**FIGURE 16.1**

The phenomenon of perspective inversion. This figure shows a wire cube viewed approximately from the direction of one corner. The phenomenon of perspective inversion makes it difficult to see which of the opposite corners of the cube is the nearer: in fact there are two stable interpretations of the cube, either of which may be perceived at any moment.

**FIGURE 16.2**

Perspective inversion for an aeroplane. Here an aeroplane (a) is silhouetted against the sky and appears as in (b). (c) shows the two planes P and Q in which the aeroplane could lie, relative to the direction D of viewing: R is the reflection plane relating the planes P and Q.

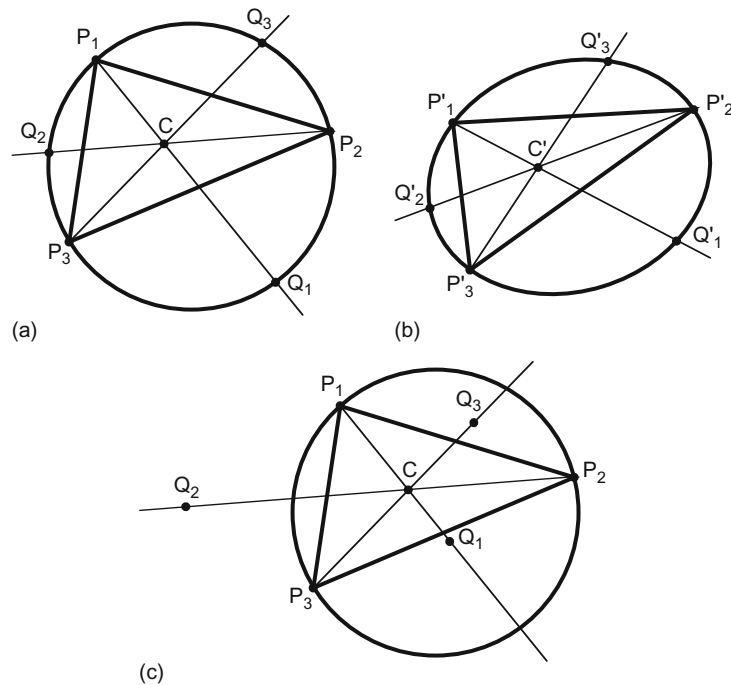
Unfortunately, a short-term study of the motion of the aeroplane will not help with interpretation in the case shown in Fig. 16.2(b). Eventually, however, the aeroplane will appear to become smaller or larger, and this will give the additional information needed to resolve the issue.

16.3 AMBIGUITY OF POSE UNDER WEAK PERSPECTIVE PROJECTION

It is instructive to examine to what extent the pose of an object can be deduced under weak perspective projection. We can reduce the above problem to a simplest case in which three points have to be located and identified. Any set of three points is coplanar, and the common plane corresponds to that of the silhouette shown in Fig. 16.2(a) (we assume here that the three points are not collinear, so that they do in fact define a plane). The problem then is to match the corresponding points on the idealized object (Fig. 16.2(a)) with those on the observed object (Fig. 16.2(b)). It is not yet completely obvious that this is possible, or that the solution is unique, even apart from the reflection operation noted earlier. It could be that more than three points will be required—especially if the scale is unknown—or it could be that there are several solutions, even if we ignore the reflection ambiguity. It will be important to see whether it is possible to distinguish the three points in the observed image.

To understand the degree of difficulty, let us briefly consider full perspective projection. In this case, any set of three noncollinear points can be mapped into any other three. This means that it may not be possible to deduce much about the original object just from this information: we will certainly not be able to deduce which point maps to which other point. However, we shall see that the situation is rather less ambiguous when viewing the object under weak perspective projection.

Perhaps the simplest approach (due to Huang et al. as recently as 1995) is to imagine a circle drawn through the original set of points P_1, P_2, P_3 (Fig. 16.3(a)). We then find the centroid C of the set of points and draw additional lines through the points, all passing through C and meeting the circle in another three points Q_1, Q_2, Q_3 (Fig. 16.3(a)). Now in common with orthographic projection, scaled orthographic projection maintains ratios of distances on the same straight line, and weak perspective projection approximates to this. Thus, the distance ratio $P_iC:CQ_i$ remains unchanged after projection. Thus, when we project the whole figure, as in Fig. 16.3(b), we find that the circle has become an ellipse, although all lines remain lines, and all linear distance ratios remain unchanged. The significance of this is as follows. When the points P'_1, P'_2, P'_3 are observed in the image, the centroid C' can be computed, as can the positions of Q'_1, Q'_2, Q'_3 . Thus, we have six points from which to deduce the position and parameters of the ellipse (in fact, five are sufficient). Once the ellipse is known, the orientation of its major axis gives the axis of rotation of the object; while the ratio of the lengths of the minor to

**FIGURE 16.3**

Determination of pose for three points viewed under weak perspective projection.

(a) shows three feature points P_1, P_2, P_3 that lie on a known type of object. The circle passing through P_1, P_2, P_3 is drawn, and lines through the points and their centroid C meet the circle in Q_1, Q_2, Q_3 . The ratios $P_iC:CQ_i$ are then deduced. (b) shows the three points observed under weak perspective as P'_1, P'_2, P'_3 , together with their centroid C' and the three points Q'_1, Q'_2, Q'_3 located using the original distance ratios. An ellipse drawn through the six points $P'_1, P'_2, P'_3, Q'_1, Q'_2, Q'_3$ can now be used to determine the orientation of the plane in which P_1, P_2, P_3 must lie, and also (from the major axis of the ellipse) the distance of viewing. (c) shows how an erroneous interpretation of the three points does not permit a circle to be drawn passing through $P_1, P_2, P_3, Q_1, Q_2, Q_3$ and hence no ellipse can be found that passes through the observed and the derived points $P'_1, P'_2, P'_3, Q'_1, Q'_2, Q'_3$.

major axes immediately gives the value of $\cos \alpha$. (Notice how the ambiguity in the sign of α comes up naturally in this calculation.) Finally, the length of the major axis of the ellipse permits the depth of the object in the scene to be deduced.

We have now shown that observing three projected points permits a unique ellipse to be computed passing through them, and when this is back-projected into a circle, the axis of rotation of the object and the angle of rotation can be deduced, but not the sign of the angle of rotation. There are two important comments to be

made about the above calculation. The first is that the three distance ratios must be stored in memory, before interpretation of the observed scene can begin. The second is that the order of the three points apparently has to be known before interpretation can be undertaken: otherwise, we will have to perform six computations in which all possible assignments of the distance ratios are tried; furthermore, it might appear from the earlier introductory remarks that several solutions are possible. While there are some instances in which feature points might be distinguishable, there are many cases when they are not (especially in 3-D situations where corner features might vary considerably when viewed from different positions). Thus, the potential ambiguity is important. However, if we can try out each of the six cases, little difficulty will generally arise. For, immediately we deduce the positions Q_1' , Q_2' , Q_3' , we will find that it is not possible in general to fit the six resulting points to an ellipse. The reason is easily seen on returning to the original circle. In that case, if the wrong distance ratios are assigned, the Q_i will clearly not lie on the circle, since the only values of the distance ratios for which the Q_i do lie on the circle are the correctly assigned ones (Fig. 16.3(c)). This means that although computation is wasted testing the incorrect assignments, there appears to be no risk of their leading to ambiguous solutions. Nevertheless, there is one contingency under which things could go wrong. Suppose the original set of points P_1 , P_2 , P_3 forms an almost perfect equilateral triangle. Then the distance ratios will be very similar, and, taking numerical inaccuracies into account, it may not be clear which ellipse provides the best and most likely fit. This mitigates against taking sets of feature points that form approximately isosceles or equilateral triangles. However, in practice more than three coplanar points will generally be used to optimize the fit, making fortuitous solutions rather unlikely.

Overall, it is fortunate that weak perspective projection requires such weak conditions for the identification of unique (to within a reflection) solutions, especially as full perspective projection demands four points before a unique solution can be found (see below). However, under weak perspective projection additional points lead to greater accuracy but no reduction in the reflection ambiguity: this is because the information content from weak perspective projection is impoverished in the lack of depth cues that could (at least in principle) resolve the ambiguity. To understand this lack of additional information from more than three points under weak perspective projection, note that each additional feature point in the same plane is predetermined once three points have been identified (here we are assuming that the model object with the correct distance ratios can be referred to).

These considerations indicate that we have two potential routes to unique location of objects from limited numbers of feature points. The first is to resort to use of noncoplanar points viewed still under weak perspective projection. The second is to use full perspective projection to view coplanar or noncoplanar sets of feature points. We shall see below that whichever of these options we take, a unique solution demands that a minimum of four feature points be located on any object.

16.4 OBTAINING UNIQUE SOLUTIONS TO THE POSE PROBLEM

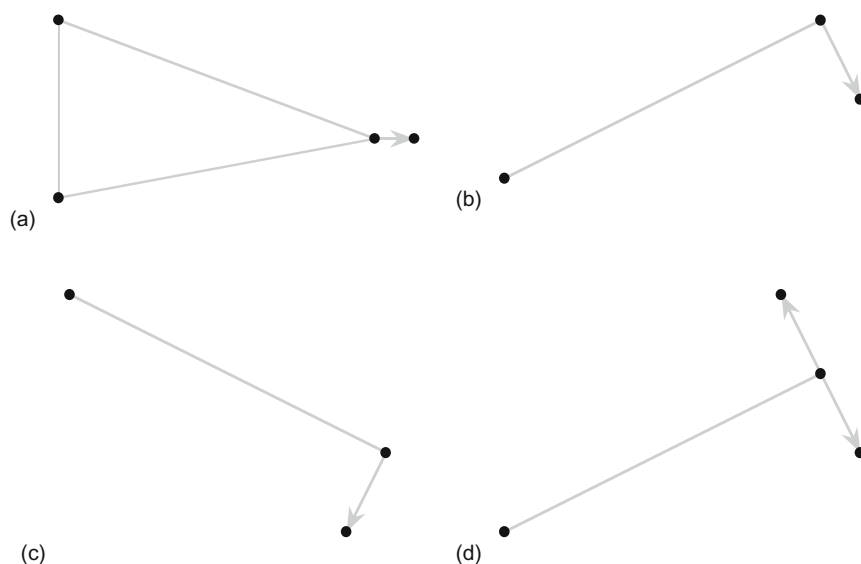
The overall situation is summarized in Table 16.1. Looking first at the case of weak perspective projection, the number of solutions only becomes finite for three or more point features. Once three points have been employed, in the coplanar case there is no further reduction in the number of solutions, since (as noted earlier) the positions of any additional points can be deduced from the existing ones. However, this does not apply when the additional points are noncoplanar since they are able to provide just the right information to eliminate any ambiguity (see Fig. 16.4). (Although this might appear to contradict what was said earlier about perspective inversion, note that we are assuming here that the body is rigid and that all its features are at *known* fixed points on it in three dimensions; hence this particular ambiguity no longer applies, except for objects with special symmetries that we shall ignore here—see Fig. 16.4(d).)

Considering next the case of full perspective projection, the number of solutions again becomes finite only for three or more point features. The lack of information provided by three point features means that four solutions are in principle possible (see the example in Fig. 16.5 and the detailed explanation in Section 16.4.1), but the number of solutions drops to one as soon as four coplanar points are employed (the correct solution can be found by making cross checks between subsets of three points, and eliminating inconsistent solutions); when the

Table 16.1 Ambiguities When Estimating Pose from Point Features

Arrangement of the Points	n	WPP	FPP
Coplanar	≤ 2	∞	∞
	3	2	4
	4	2	1
	5	2	1
	≥ 6	2	1
Noncoplanar	≤ 2	∞	∞
	3	2	4
	4	1	2
	5	1	2
	≥ 6	1	1

This table summarizes the numbers of solutions that will be obtained when estimating the pose of a rigid object from point features located in a single image. It is assumed that n point features are detected and identified correctly and in the correct order. The columns WPP and FPP signify weak perspective projection and full perspective projection, respectively. The upper half of the table applies when all n points are coplanar; the lower half of the table applies when the n points are noncoplanar. Note that when $n \leq 3$, the results strictly apply only in the coplanar case. However, the top two lines in the lower half of the table are retained for easy comparison.

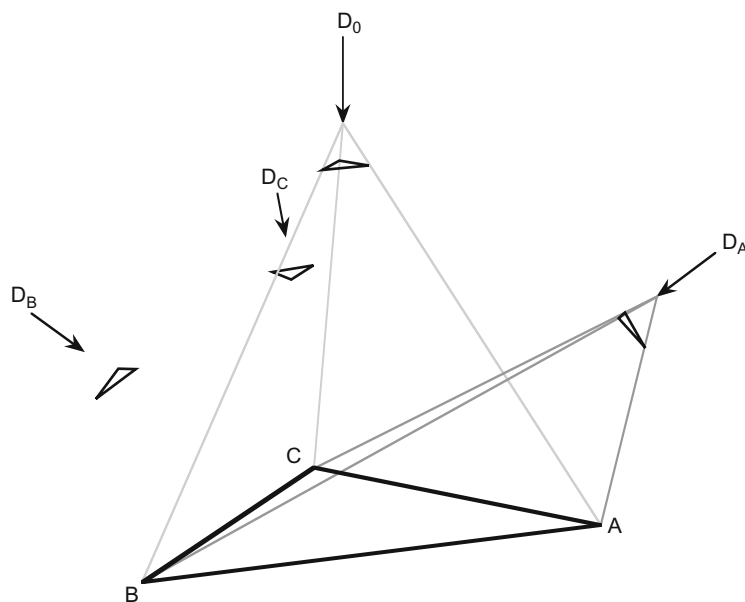
**FIGURE 16.4**

Determination of pose for four points viewed under weak perspective projection. (a) shows an object containing four noncoplanar points, as seen under weak perspective projection. (b) shows a side view of the object. If the first three points (connected by nonarrowed gray lines) were viewed alone, perspective inversion would give rise to a second interpretation (c). However, the fourth point gives additional information about the pose that permits only one overall interpretation. This would not be the case for an object containing an additional symmetry as in (d), since its reflection would be identical to the original view (not shown).

points are noncoplanar, it is only when six or more points are employed that there is sufficient information to unambiguously determine the pose: there is necessarily no ambiguity with six or more points, as all eleven camera calibration parameters can be deduced from the twelve linear equations that then arise (see Chapter 18). Correspondingly, it is deduced that five noncoplanar points will in general be *insufficient* for all eleven parameters to be deduced, so there will still be some ambiguity in this case.

Next, it should be questioned why the coplanar case is at first ($n = 3$) better⁴ under weak perspective projection and then ($n > 3$) better under full perspective projection, while the noncoplanar case is always better, or as good, under weak perspective projection. The reason must be that intrinsically full perspective projection provides more detailed information, but is frustrated by lack of data when there are relatively few points: however, the exact stage at which the additional

⁴In this context “better” means less ambiguous, and leading to fewer solutions.

**FIGURE 16.5**

Ambiguity for three points viewed under full perspective projection. Under full perspective projection, the camera sees three points A, B, C as three directions in space, and this can lead to four-fold ambiguity in interpreting a known object. The figure shows the four possible viewing directions and centers of projection of the camera (indicated by the directions and tips of the bold arrows): in each case the image at each camera is indicated by a small triangle. D_A , D_B , D_C correspond approximately to views from the general directions of A, B, C, respectively.

information becomes available is different in the coplanar and noncoplanar cases. In this respect, it is important to note that when coplanar points are being observed under weak perspective projection, there is never enough information to eliminate the ambiguity.

It should be emphasized that the above discussion assumes that the correspondences between object and image features are all known, i.e., n point features are detected and identified correctly and in the correct order. If this is not so, the number of possible solutions could increase substantially, considering the number of possible permutations of quite small numbers of points. This makes it attractive to use the minimum number of features for ascertaining the most probable match (Horaud et al., 1989). Other workers have used heuristics to help reduce the number of possibilities. For example, Tan (1995) used a simple compactness measure (see Section 9.7) to determine which is the most likely geometric solution: extreme obliqueness is perhaps unlikely, and the most likely solution is taken to be the one with highest compactness value. This idea follows on from the extremum

principle of Brady and Yuille (1984), which states that the most probable solutions are those nearest to extrema of relevant (e.g., rotation) parameters.⁵ In this context, note that coplanar points viewed under weak or full perspective projection always appear in the same cyclic order: this is not trivial to check given the possible distortions of an object, although if a convex polygon can be drawn through the points, the cyclic order around its boundary will not change on projection.⁶ However, for noncoplanar points, the pattern of the perceived points can re-order itself almost randomly: this means that a considerably greater number of permutations of the points have to be considered for noncoplanar points than for coplanar points.

Finally, note that the above discussion has concentrated on the existence and uniqueness of solutions to the pose problem. The stability of the solutions has not so far been discussed. However, the concept of stability gives a totally different dimension to the data presented in Table 16.1. In particular, noncoplanar points tend to give more stable solutions to the pose problem. For example, if the plane containing a set of coplanar points is viewed almost head-on ($\alpha \approx 0$), there will be very little information on the exact orientation of the plane, because the changes in lateral displacement of the points will vary as $\cos \alpha$ (see Section 16.2) and there will be no linear term in the Taylor expansion of the orientation dependence.

16.4.1 Solution of the Three-Point Problem

Figure 16.5 showed how four solutions can arise when three point features are viewed under full perspective projection. Here, we briefly explore this situation by considering the relevant equations. Figure 16.5 shows that the camera sees the points as three image points representing three directions in space. This means that we can compute the angles α, β, γ between these three directions. If the distances between the three points A, B, C on the object are the known values D_{AB}, D_{BC}, D_{CA} , we can now apply the cosine rule in an attempt to determine the distances R_A, R_B, R_C of the feature points from the center of projection:

$$D_{BC}^2 = R_B^2 + R_C^2 - 2R_BR_C \cos \alpha \quad (16.1)$$

$$D_{CA}^2 = R_C^2 + R_A^2 - 2R_CR_A \cos \beta \quad (16.2)$$

$$D_{AB}^2 = R_A^2 + R_B^2 - 2R_AR_B \cos \gamma \quad (16.3)$$

Eliminating any two of the variables R_A, R_B, R_C yields an eighth degree equation in the other variable, indicating that eight solutions to the system of equations could be available (Fischler and Bolles, 1981). However, the above cosine rule equations contain only constants and second degree terms: hence, for every positive solution there is another solution that differs only by a sign change in all the

⁵Perhaps the simplest way of understanding this principle is obtained by considering a pendulum, whose extreme positions are also its most probable. However, in this case the extremum occurs when the angle α (see Fig. 16.1) is close to zero.

⁶The reason for this is that planar convexity is an invariant of projection.

variables. These solutions correspond to inversion through the center of projection and are hence unrealizable. Thus, there are at most four realizable solutions to the system of equations. In fact, we can quickly demonstrate that there may sometimes be fewer than four solutions: since in some cases, for one or more of the “flipped” positions shown in Fig. 16.5, one of the features could be on the negative side of the center of projection, and hence would be unrealizable.

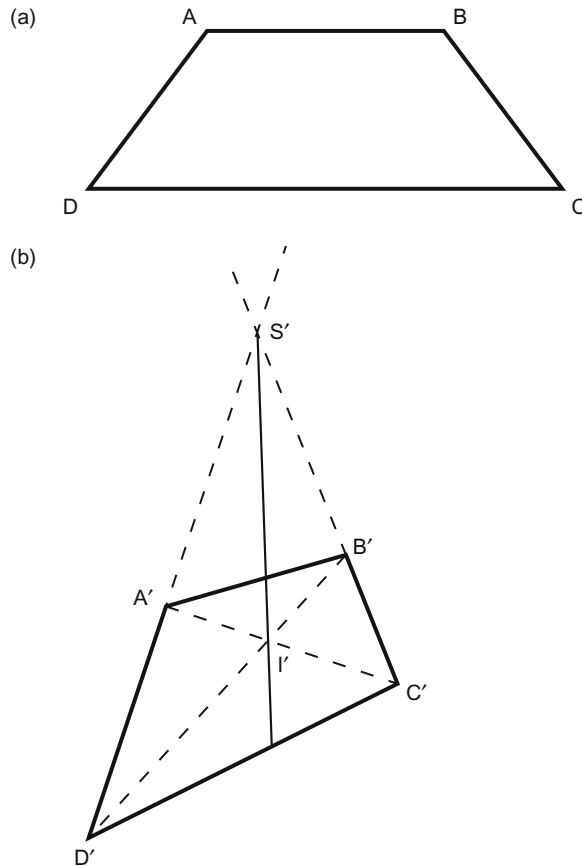
Before leaving this topic, note that the homogeneity of Eqs. (16.1)–(16.3) implies that observation of the angles α, β, γ permits the orientation of the object to be estimated independently of any knowledge of its scale: in fact, estimation of scale depends directly on estimation of range, and *vice versa*. Thus, knowledge of just one range parameter (e.g., R_A) will permit the scale of the object to be deduced. Alternatively, knowledge of its area will permit the remaining parameters to be deduced. This concept provides a slight generalization of the main results of Sections 16.2 and 16.3, which generally start with the assumption that all the dimensions of the object are known.

16.4.2 Using Symmetric Trapezia for Estimating Pose

One more example will be of interest here. That is the case of four points arranged at the corners of a symmetric trapezium (Tan, 1995). When viewed under weak perspective projection, the mid-points of the parallel sides are easily measured, but under full perspective projection midpoints do not remain midpoints, so the axis of symmetry cannot be obtained in this way. However, producing the skewed sides to meet at S' and forming the intersection I' of the diagonals permit the axis of symmetry to be located as the line $I'S'$ (Fig. 16.6). Thus, we now have not four points but six to describe the perspective view of the trapezium. What is more important is that the axis of symmetry has been located and this is known to be perpendicular to the parallel sides of the trapezium. This is a great help in making the mathematics more tractable and in obtaining solutions quickly so that, e.g., object motion can be tracked in real time. Again, this is a case where object orientation can be deduced straightforwardly, even when the situation is one of strong perspective, and even when the size of the object is unknown. This result is a generalization from that of Haralick (1989) who noted that a single view of a rectangle of unknown size is sufficient to determine its pose. In either case, the range of the object can be found if its area is known, or its size can be deduced if a single range value can be found from other data (see also Section 16.4.1).

16.5 CONCLUDING REMARKS

This chapter has aimed to cover certain aspects of 3-D vision that were not studied in depth in the previous chapter. In particular, it was worth investigating the topic of perspective inversion in some detail and exploring how it was affected by the method of projection. Orthographic projection, scaled orthographic

**FIGURE 16.6**

Trapezium viewed under full perspective projection. (a) shows a symmetrical trapezium, and (b) shows how it appears when viewed under full perspective projection. In spite of the fact that midpoints do not project into midpoints under perspective projection, the two points S' and I' on the symmetry axis can be located unambiguously as the intersection of two nonparallel sides and two diagonals, respectively. This gives six points (from which the two midpoints on the symmetry axis can be deduced if necessary), which is sufficient to compute the pose of the object, albeit with a single ambiguity of interpretation (see text).

projection, weak perspective projection, and full perspective projection were considered, and the numbers of object points that would lead to correct or ambiguous interpretations were analyzed. It was found that scaled orthographic projection and its approximation, weak perspective projection, led to straightforward interpretation when four or more noncoplanar points were considered,

although the perspective inversion ambiguity remained when all the points were coplanar. This latter ambiguity was resolved with four or more points viewed under full perspective projection. However, in the noncoplanar case, some ambiguity remained until six points were being viewed. The key to understanding the situation is the fact that full perspective projection makes the situation more complex, although at the same time it provides more information, ultimately, to resolve the ambiguity.

Additional problems were found to arise when the points being viewed are indistinguishable, and then a good many solutions may have to be tried before a minimally ambiguous result is obtained. With coplanar points fewer possibilities arise and this leads to less computational complexity: the key to success here is the natural ordering that can arise for points in a plane—e.g., when they form a convex set that can be ordered uniquely around a bounding polygon. In this context, the role that can be played by the extremum principle of Brady and Yuille (1984) in reducing the number of solutions is significant (for further insight on the topic, see Horaud and Brady, 1988).

It is of great relevance to devise methods for rapid interpretation in real-time applications. To achieve this, it is important to work with a minimal set of points and to obtain analytic solutions that move directly to solutions without computationally expensive iterative procedures: for example, Horaud et al. (1989) found an analytic solution for the perspective four-point problem that works both in the general noncoplanar case and in the planar case. Other low computation methods are still being developed, as with pose determination for symmetrical trapezia (Tan, 1995). It should also be noted that understanding is still advancing, as demonstrated by Huang et al.'s (1995) neat geometrical solution to the pose determination problem for three points viewed under weak perspective projection.

This chapter has covered a specific 3-D recognition problem. Chapter 17 covers another—that of invariants, which provides speedy and convenient means of by-passing the difficulties associated with full perspective projection. Chapter 18 aims to finalize the study of 3-D vision by showing how camera calibration can be achieved or, to some extent, circumvented.

Perspective makes interpretation of images of 3-D scenes intrinsically difficult. However, this chapter has demonstrated that “weak perspective” views of distant objects are much simplified, so objects are commonly located using fewer features: for planar objects a pose ambiguity remains, although it is eliminated under full perspective.

16.6 BIBLIOGRAPHICAL AND HISTORICAL NOTES

The development of solutions to the so-called *perspective n -point (PnP) problem* (finding the pose of objects from n features under various forms of perspective) has been proceeding for more than two decades, and is by no means complete.

Fischler and Bolles summarized the situation as they saw it in 1981, and several new algorithms were described by them. However, they did not discuss pose determination under weak perspective, and perhaps surprisingly, considering its reduced complexity, this has subsequently been the subject of much research (e.g., Alter, 1994; Huang et al., 1995). Horaud et al. (1989) discussed the problem of finding rapid solutions to the PnP problem by reducing n as far as possible: they also obtained an analytic solution for the case $n = 4$ which should help considerably with real-time implementations. Their solution is related to Horaud's earlier (1987) corner interpretation scheme—described in Section 15.10, while Haralick et al. (1984) provided useful basic theory for matching wire frame objects.

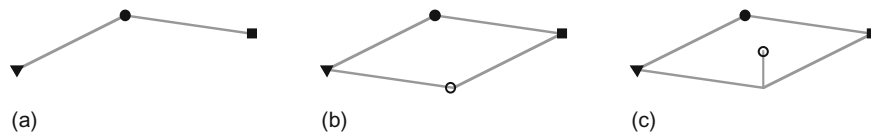
In a later paper, Liu and Wong (1999) described an algorithm for pose estimation from four corresponding points under FPP when the points are not coplanar. Strictly, according to Table 16.1, this will lead to an ambiguity. However, Liu and Wong made the point “that the possibility for the occurrence of multiple solutions in the perspective 4-point problem is much smaller than that in the perspective 3-point problem,” so that “using a 4-point model is much more reliable than using a 3-point model,” and they actually only claim “good results.” Also, much of the emphasis of the paper is on errors and reliability. Hence, it seems that it is the *scope* for making errors in the sense of misinterpreting the situation that is significantly reduced. Added to this, Liu and Wong's (1999) work involves tracking a known object within a somewhat restricted region of space: this must again cut down the scope for error considerably. Hence it is not clear that their work violates the relevant (FPP; noncoplanar; $n = 4$)⁷ entry in Table 16.1, rather than merely making it *unlikely* that a real ambiguity will arise.

Between them, Faugeras (1993), Hartley and Zisserman (2000), Faugeras and Luong (2001), and Forsyth and Ponce (2003) provide good coverage of the whole area of 3-D vision; for an interesting viewpoint on the subject, with particular emphasis on pose refinement, see Sullivan (1992). For further references on specific aspects of 3-D vision, see Sections 15.13, 17.13, and 18.16. (Section 19.10 gives references on motion, but also covers aspects of 3-D vision.)

16.6.1 More Recent Developments

Xu et al. (2008) present a new method for tackling the PnP problem. The linear method for the case of four coplanar points is extended to find coarse solutions for the general P3P problem. Once all the accurate solutions for the P3P problem have been found, the algorithm is applied to the general PnP problem. Solution stability issues and possible ambiguities are investigated and experiments are performed to verify the effectiveness of the proposed method. Interestingly, the case of four coplanar points has to be divided into two mutually exclusive cases, in which one point lies, or does not lie, within the triangle presented by the other three points. Lepetit et al. (2008) propose an $O(n)$

⁷See Fischler and Bolles (1981) for the original evidence for this particular ambiguity.

**FIGURE 16.7**

In this diagram the gray edges are construction lines, *not* parts of the objects. (a) and (b) are completely planar objects.

noniterative solution to the PnP problem that is far faster and more accurate than earlier methods and also more stable. The central idea is to express the 3-D points as a weighted sum of four virtual control points and to solve this case in terms of the coordinates. (The computational load of previous methods can be as high as $O(n^5)$ or even $O(n^8)$.)

16.7 PROBLEMS

1. Draw up a complete table of pose ambiguities that arise for *weak* perspective projection, for various numbers of object points identified in the image. Your answer should cover both coplanar points and noncoplanar points, and should make clear in each case how much ambiguity would remain in the limit of an infinite number of object points being seen. Give justification for your results.
2. Distinguish between *full perspective projection* and *weak perspective projection*. Explain how each of these projections presents oblique views of the following real objects: (i) straight lines, (ii) several concurrent lines (i.e., lines meeting in a single point), (iii) parallel lines, (iv) midpoints of lines, (v) tangents to curves, (vi) circles whose centers are marked with a dot. Give justification for your results.
3. Explain each of the following: (i) Why weak perspective projection leads to an ambiguity in viewing an object such as that in Fig. 16.7(a). (ii) Why the ambiguity doesn't disappear for the case of Fig. 16.7(b). (iii) Why the ambiguity *does* disappear in the case of Fig. 16.7(c), if the true nature of the object is known. (iv) Why the ambiguity doesn't occur in the case of Fig. 16.7(b) viewed under full perspective projection. In the last case, illustrate your answer by means of a sketch.