10

3D Reconstruction of Cameras and Structure

This chapter describes how and to what extent the spatial layout of a scene and the cameras can be recovered from two views. Suppose that a set of image correspondences $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ are given. It is assumed that these correspondences come from a set of 3D points $\mathbf{X}_i$, which are unknown. Similarly, the position, orientation and calibration of the cameras are not known. The reconstruction task is to find the camera matrices $\mathtt{P}$ and $\mathtt{P}'$, as well as the 3D points $\mathbf{X}_i$ such that

$$\mathbf{x}_i = \mathtt{P}\mathbf{X}_i \quad \mathbf{x}'_i = \mathtt{P}'\mathbf{X}_i \quad \text{for all } i.$$

Given too few points, this task is not possible. However, if there are sufficiently many point correspondences to allow the fundamental matrix to be computed uniquely, then the scene may be reconstructed up to a projective ambiguity. This is a very significant result, and one of the major achievements of the uncalibrated approach.

The ambiguity in the reconstruction may be reduced if additional information is supplied on the cameras or scene. We describe a two-stage approach where the ambiguity is first reduced to affine, and second to metric; each stage requiring information of the appropriate class.

10.1 Outline of reconstruction method

We describe a method for reconstruction from two views as follows.

(i) Compute the fundamental matrix from point correspondences.
(ii) Compute the camera matrices from the fundamental matrix.
(iii) For each point correspondence $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$, compute the point in space that projects to these two image points.

Many variants on this method are possible. For instance, if the cameras are calibrated, then one will compute the essential matrix instead of the fundamental matrix. Furthermore, one may use information about the motion of the camera, scene constraints or partial camera calibration to obtain refinements of the reconstruction.

Each of the steps of this reconstruction method will be discussed briefly in the following paragraphs. The method described is no more than a conceptual approach to reconstruction. The reader is warned not to implement a reconstruction method based solely on the description given in this section. For real images where measurements

Downloaded from https://www.cambridge.org/core. University of Illinois at Urbana - Champaign Library, on 22 Sep 2018 at 15:36:07, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811685.015

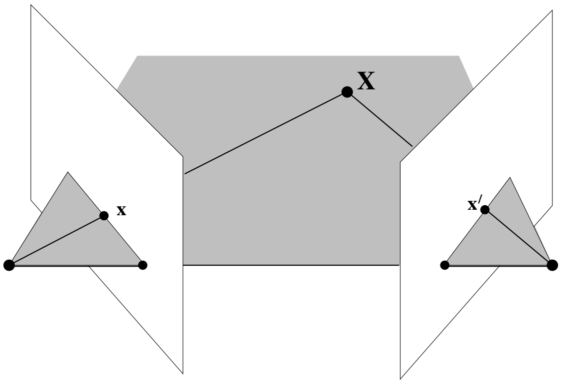

Fig. 10.1. **Triangulation.** *The image points* $\mathbf{x}$ *and* $\mathbf{x}'$ *back project to rays. If the epipolar constraint* $\mathbf{x}'^{\mathsf{T}}\mathtt{F}\mathbf{x} = 0$ *is satisfied, then these two rays lie in a plane, and so intersect in a point* $\mathbf{X}$ *in 3-space.*

are "noisy" preferred methods for reconstruction, based on this general outline, are described in chapter 11 and chapter 12.

Computation of the fundamental matrix. Given a set of correspondences $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ in two images the fundamental matrix $\mathtt{F}$ satisfies the condition $\mathbf{x}'_i\mathtt{F}\mathbf{x}_i = 0$ for all i. With the $\mathbf{x}_i$ and $\mathbf{x}'_i$ known, this equation is linear in the (unknown) entries of the matrix $\mathtt{F}$. In fact, each point correspondence generates one linear equation in the entries of $\mathtt{F}$. Given at least 8 point correspondences it is possible to solve linearly for the entries of $\mathtt{F}$ up to scale (a non-linear solution is available for 7 point correspondences). With more than 8 equations a least-squares solution is found. This is the general principle of a method for computing the fundamental matrix.

Recommended methods of computing the fundamental matrix from a set of point correspondences will be described later in chapter 11.

Computation of the camera matrices. A pair of camera matrices $\mathtt{P}$ and $\mathtt{P}'$ corresponding to the fundamental matrix $\mathtt{F}$ are easily computed using the direct formula in result 9.14.

Triangulation. Given the camera matrices $\mathtt{P}$ and $\mathtt{P}'$, let $\mathbf{x}$ and $\mathbf{x}'$ be two points in the two images that satisfy the epipolar constraint, $\mathbf{x}'^{\mathsf{T}}\mathtt{F}\mathbf{x} = 0$. As shown in chapter 9 this constraint may be interpreted geometrically in terms of the rays in space corresponding to the two image points. In particular it means that $\mathbf{x}'$ lies on the epipolar line $\mathtt{F}\mathbf{x}$. In turn this means that the two rays back-projected from image points $\mathbf{x}$ and $\mathbf{x}'$ lie in a common epipolar plane, that is, a plane passing through the two camera centres. Since the two rays lie in a plane, they will intersect in some point. This point $\mathbf{X}$ projects via the two cameras to the points $\mathbf{x}$ and $\mathbf{x}'$ in the two images. This is illustrated in figure 10.1.

The only points in 3-space that cannot be determined from their images are points on the baseline between the two cameras. In this case the back-projected rays are collinear (both being equal to the baseline) and intersect along their whole length. Thus, the point

Downloaded from https://www.cambridge.org/core. University of Illinois at Urbana - Champaign Library, on 22 Sep 2018 at 15:36:07, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811685.015

$\mathbf{X}$ cannot be uniquely determined. Points on the baseline project to the epipoles in both images.

Numerically stable methods of actually determining the point $\mathbf{X}$ at the intersection of the two rays back-projected from $\mathbf{x}$ and $\mathbf{x}'$ will be described later in chapter 12.

10.2 Reconstruction ambiguity

In this section we discuss the inherent ambiguities involved in reconstruction of a scene from point correspondences. This topic will be discussed in a general context, without reference to a specific method of carrying out the reconstruction.

Without some knowledge of a scene's placement with respect to a 3D coordinate frame, it is generally not possible to reconstruct the absolute position or orientation of a scene from a pair of views (or in fact from any number of views). This is true independently of any knowledge which may be available about the internal parameters of the cameras, or their relative placement. For instance the exact latitude and longitude of the scene in figure 9.8(*p*248) (or any scene) cannot be computed, nor is it possible to determine whether the corridor runs north-south or east-west. This may be expressed by saying that the scene is determined at best up to a Euclidean transformation (rotation and translation) with respect to the world frame.

Only slightly less obvious is the fact that the overall scale of the scene cannot be determined. Considering figure 9.8(*p*248) once more, it is impossible based on the images alone to determine the width of the corridor. It may be two metres, one metre. It is even possible that this is an image of a doll's house and the corridor is 10 cm wide. Our common experience leads us to expect that ceilings are approximately 3m from the floor, which allows us to perceive the real scale of the scene. This extra information is an example of subsidiary knowledge of the scene not derived from image measurements. Without such knowledge therefore the scene is determined by the image only up to a similarity transformation (rotation, translation and scaling).

To give a mathematical basis to this observation, let $\mathbf{X}_i$ be a set of points and $\mathtt{P}, \mathtt{P}'$ be a pair of cameras projecting $\mathbf{X}_i$ to image points $\mathbf{x}_i$ and $\mathbf{x}'_i$. The points $\mathbf{X}_i$ and the camera pair constitute a reconstruction of the scene from the image correspondences. Now let

$$\mathtt{H}_\mathtt{S} = \begin{bmatrix} \mathtt{R} & \mathbf{t} \\ \mathbf{0}^\mathsf{T} & \lambda \end{bmatrix}$$

be any similarity transformation: $\mathtt{R}$ is a rotation, $\mathbf{t}$ a translation and λ^{-1} represents overall scaling. Replacing each point $\mathbf{X}_i$ by $\mathtt{H}_\mathtt{S}\mathbf{X}_i$ and cameras $\mathtt{P}$ and $\mathtt{P}'$ by $\mathtt{P}\mathtt{H}_\mathtt{S}^{-1}$ and $\mathtt{P}'\mathtt{H}_\mathtt{S}^{-1}$ respectively does not change the observed image points, since $\mathtt{P}\mathbf{X}_i = (\mathtt{P}\mathtt{H}_\mathtt{S}^{-1})(\mathtt{H}_\mathtt{S}\mathbf{X}_i)$. Furthermore, if $\mathtt{P}$ is decomposed as $\mathtt{P} = \mathtt{K}[\mathtt{R}_\mathtt{P} \mid \mathbf{t}_\mathtt{P}]$, then one computes

$$\mathtt{P}\mathtt{H}_\mathtt{S}^{-1} = \mathtt{K}[\mathtt{R}_\mathtt{P}\mathtt{R}^{-1} \mid \mathbf{t}']$$

for some $\mathbf{t}'$ that we do not need to compute more exactly. This result shows that multiplying by $\mathtt{H}_\mathtt{S}^{-1}$ does not change the calibration matrix of $\mathtt{P}$. Consequently this ambiguity of reconstruction exists even for calibrated cameras. It was shown by Longuet-Higgins

Downloaded from https://www.cambridge.org/core. University of Illinois at Urbana - Champaign Library, on 22 Sep 2018 at 15:36:07, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811685.015

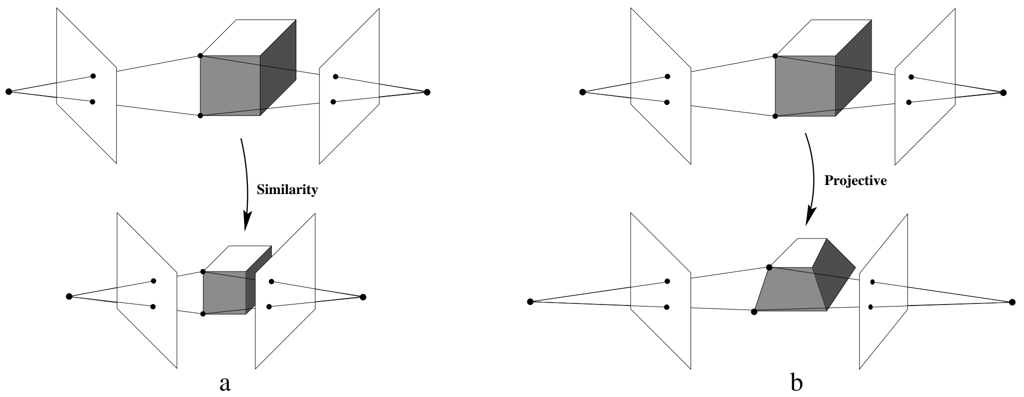

Fig. 10.2. **Reconstruction ambiguity.** *(a) If the cameras are calibrated then any reconstruction must respect the angle between rays measured in the image. A similarity transformation of the structure and camera positions does not change the measured angle. The angle between rays and the baseline (epipoles) is also unchanged. (b) If the cameras are uncalibrated then reconstructions must only respect the image points (the intersection of the rays with the image plane). A projective transformation of the structure and camera positions does not change the measured points, although the angle between rays is altered. The epipoles are also unchanged (intersection with baseline).*

([LonguetHiggins-81]) that for calibrated cameras, this is the only ambiguity of reconstruction. Thus for calibrated cameras, reconstruction is possible *up to a similarity transformation.* This is illustrated in figure 10.2a.

Projective ambiguity. If nothing is known of the calibration of either camera, nor the placement of one camera with respect to the other, then the ambiguity of reconstruction is expressed by an arbitrary projective transformation. In particular, if $\mathtt{H}$ is any 4×4 invertible matrix, representing a projective transformation of $\mathbb{P}^3$, then replacing points $\mathbf{X}_i$ by $\mathtt{H}\mathbf{X}_i$ and matrices $\mathtt{P}$ and $\mathtt{P}'$ by $\mathtt{PH}^{-1}$ and $\mathtt{P}'\mathtt{H}^{-1}$ (as in the previous paragraph) does not change the image points. This shows that the points $\mathbf{X}_i$ and the cameras can be determined at best only up to a projective transformation. It is an important result, proved in this chapter (section 10.3), that this is the only ambiguity in the reconstruction of points from two images. Thus reconstruction from uncalibrated cameras is possible *up to a projective transformation.* This is illustrated in figure 10.2b.

Other types of reconstruction ambiguity result from certain assumptions on the types of motion, or partial knowledge of the cameras. For instance,

(i) If the two cameras are related via a translational motion, without change of calibration, then reconstruction is possible up to an affine transformation.
(ii) If the two cameras are calibrated apart from their focal lengths, then reconstruction is still possible up to a similarity transformation.

These two cases will be considered later in section 10.4.1 and example 19.8(*p*472), respectively.

Terminology. In any reconstruction problem derived from real data, consisting of point correspondences $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$, there exists a **true** reconstruction consisting of the actual points $\bar{\mathbf{X}}_i$ and actual cameras $\bar{\mathtt{P}}, \bar{\mathtt{P}}'$ that generated the measured observations. The

Downloaded from https://www.cambridge.org/core. University of Illinois at Urbana - Champaign Library, on 22 Sep 2018 at 15:36:07, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811685.015

reconstructed point set $\mathbf{X}_i$ and cameras differ from the true reconstruction by a transformation belonging to a given class or group (for instance a similarity, projective or affine transformation). One speaks of *projective reconstruction*, *affine reconstruction*, *similarity reconstruction*, and so on, to indicate the type of transformation involved. However, the term *metric reconstruction* is normally used in preference to *similarity reconstruction*, being identical in meaning. The term indicates that metric properties, such as angles between lines and ratios of lengths, can be measured on the reconstruction and have their veridical values (since these are similarity invariants). In addition, the term *Euclidean reconstruction* is frequently used in the published literature to mean the same thing as a similarity or metric reconstruction, since true Euclidean reconstruction (including determination of overall scale) is not possible without extraneous information.

10.3 The projective reconstruction theorem

In this section the basic theorem of projective reconstruction from uncalibrated cameras is proved. Informally, the theorem may be stated as follows.

- *If a set of point correspondences in two views determine the fundamental matrix uniquely, then the scene and cameras may be reconstructed from these correspondences alone, and any two such reconstructions from these correspondences are projectively equivalent.*

Points lying on the line joining the two camera centres must be excluded, since such points cannot be reconstructed uniquely even if the camera matrices are determined. The formal statement is:

Theorem 10.1 (Projective reconstruction theorem). *Suppose that* $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ *is a set of correspondences between points in two images and that the fundamental matrix* $\mathtt{F}$ *is uniquely determined by the condition* $\mathbf{x}'^{\mathsf{T}}_i \mathtt{F}\mathbf{x}_i = 0$ *for all* i. *Let* $(\mathtt{P}_1, \mathtt{P}'_1, \{\mathbf{X}_{1i}\})$ *and* $(\mathtt{P}_2, \mathtt{P}'_2, \{\mathbf{X}_{2i}\})$ *be two reconstructions of the correspondences* $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$. *Then there exists a non-singular matrix* $\mathtt{H}$ *such that* $\mathtt{P}_2 = \mathtt{P}_1\mathtt{H}^{-1}$, $\mathtt{P}'_2 = \mathtt{P}'_1\mathtt{H}^{-1}$ *and* $\mathbf{X}_{2i} = \mathtt{H}\mathbf{X}_{1i}$ *for all* i, **except** *for those* i *such that* $\mathtt{F}\mathbf{x}_i = \mathbf{x}'^{\mathsf{T}}_i\mathtt{F} = \mathbf{0}$.

Proof. Since the fundamental matrix is uniquely determined by the point correspondences, one deduces that $\mathtt{F}$ is the fundamental matrix corresponding to the camera pair $(\mathtt{P}_1, \mathtt{P}'_1)$ and also to $(\mathtt{P}_2, \mathtt{P}'_2)$. According to theorem 9.10(*p*254) there is a projective transformation $\mathtt{H}$ such that $\mathtt{P}_2 = \mathtt{P}_1\mathtt{H}^{-1}$ and $\mathtt{P}'_2 = \mathtt{P}'_1\mathtt{H}^{-1}$ as required.
As for the points, one observes that $\mathtt{P}_2(\mathtt{H}\mathbf{X}_{1i}) = \mathtt{P}_1\mathtt{H}^{-1}\mathtt{H}\mathbf{X}_{1i} = \mathtt{P}_1\mathbf{X}_{1i} = \mathbf{x}_i$. On the other hand $\mathtt{P}_2\mathbf{X}_{2i} = \mathbf{x}_i$, so $\mathtt{P}_2(\mathtt{H}\mathbf{X}_{1i}) = \mathtt{P}_2\mathbf{X}_{2i}$. Thus both $\mathtt{H}\mathbf{X}_{1i}$ and $\mathbf{X}_{2i}$ map to the same point $\mathbf{x}_i$ under the action of the camera $\mathtt{P}_2$. It follows that both $\mathtt{H}\mathbf{X}_{1i}$ and $\mathbf{X}_{2i}$ lie on the same ray through the camera centre of $\mathtt{P}_2$. Similarly, it may be deduced that these two points lie on the same ray through the camera centre of $\mathtt{P}'_2$. There are two possibilities: either $\mathbf{X}_{2i} = \mathtt{H}\mathbf{X}_{1i}$ as required, or they are distinct points lying on the line joining the two camera centres. In this latter case, the image points $\mathbf{x}_i$ and $\mathbf{x}'_i$ coincide with the epipoles in the two images, and so $\mathtt{F}\mathbf{x}_i = \mathbf{x}'^{\mathsf{T}}_i\mathtt{F} = \mathbf{0}$. $\square$

Downloaded from https://www.cambridge.org/core. University of Illinois at Urbana - Champaign Library, on 22 Sep 2018 at 15:36:07, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811685.015

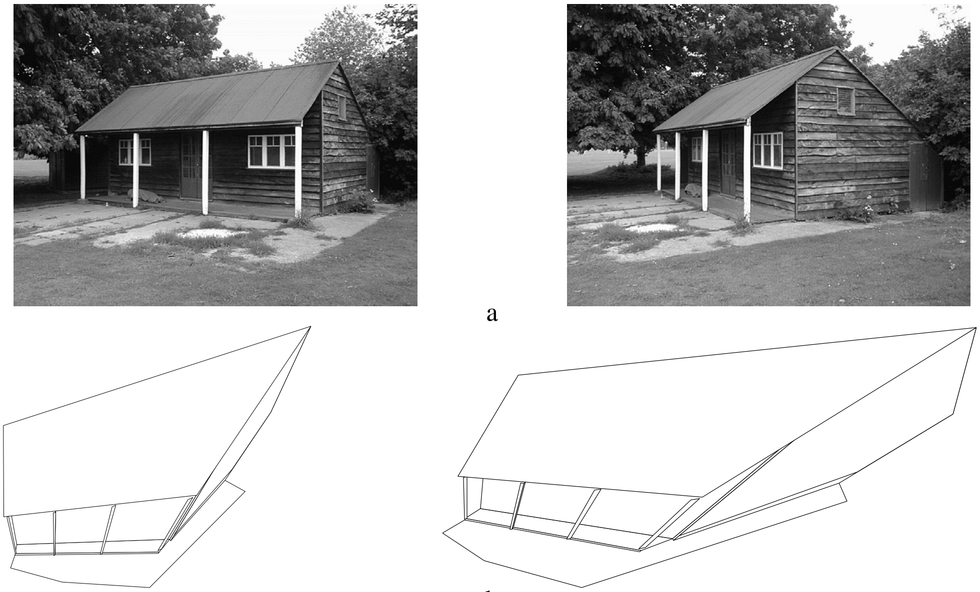

Fig. 10.3. **Projective reconstruction.** *(a) Original image pair. (b) 2 views of a 3D projective reconstruction of the scene. The reconstruction requires no information about the camera matrices, or information about the scene geometry. The fundamental matrix* $\mathtt{F}$ *is computed from point correspondences between the images, camera matrices are retrieved from* $\mathtt{F}$*, and then 3D points are computed by triangulation from the correspondences. The lines of the wireframe link the computed 3D points.*

This is an enormously significant result, since it implies that one may compute a projective reconstruction of a scene from two views based on image correspondences alone, without knowing anything about the calibration or pose of the two cameras involved. In particular the true reconstruction is within a projective transformation of the projective reconstruction. Figure 10.3 shows an example of 3D structure computed as part of a projective reconstruction from two images.

In more detail suppose the true Euclidean reconstruction is $(\mathtt{P}_{\mathrm{E}}, \mathtt{P}'_{\mathrm{E}}, \{\mathbf{X}_{\mathrm{E}i}\})$ and the projective reconstruction is $(\mathtt{P}, \mathtt{P}', \{\mathbf{X}_i\})$, then the reconstructions are related by a non-singular matrix $\mathtt{H}$ such that

$$\mathtt{P}_{\mathrm{E}} = \mathtt{P}\mathtt{H}^{-1}, \ \ \mathtt{P}'_{\mathrm{E}} = \mathtt{P}'\mathtt{H}^{-1}, \text{ and } \mathbf{X}_{\mathrm{E}i} = \mathtt{H}\mathbf{X}_i \tag{10.1}$$

where $\mathtt{H}$ is a 4×4 homography matrix which is unknown but the same for all points.

For some applications projective reconstruction is all that is required. For example, questions such as “at what point does a line intersect a plane?”, “what is the mapping between two views induced by particular surfaces, such as a plane or quadric?” can be dealt with directly from the projective reconstruction. Furthermore it will be seen in the sequel that obtaining a projective reconstruction of a scene is the first step towards affine or metric reconstruction.

10.4 Stratified reconstruction

The “stratified” approach to reconstruction is to begin with a projective reconstruction and then to refine it progressively to an affine and finally a metric reconstruction, if

Downloaded from https://www.cambridge.org/core. University of Illinois at Urbana - Champaign Library, on 22 Sep 2018 at 15:36:07, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811685.015

possible. Of course, as has just been seen, affine and metric reconstruction are not possible without further information either about the scene, the motion or the camera calibration.

10.4.1 The step to affine reconstruction

The essence of affine reconstruction is to locate the plane at infinity by some means, since this knowledge is equivalent to an affine reconstruction. This equivalence is explained in the 2D case in section 2.7(*p*47). To see this equivalence for reconstruction, suppose we have determined a projective reconstruction of a scene, consisting of a triple $(\mathtt{P}, \mathtt{P}', \{\mathbf{X}_i\})$. Suppose further that by some means a certain plane $\boldsymbol{\pi}$ has been identified as the true plane at infinity. The plane $\boldsymbol{\pi}$ is expressed as a 4-vector in the coordinate frame of the projective reconstruction. In the true reconstruction, $\boldsymbol{\pi}$ has coordinates $(0,0,0,1)^\mathsf{T}$, and we may find a projective transformation that maps $\boldsymbol{\pi}$ to $(0,0,0,1)^\mathsf{T}$. Considering the way a projective transformation acts on planes, we want to find $\mathtt{H}$ such that $\mathtt{H}^{-\mathsf{T}}\boldsymbol{\pi} = (0,0,0,1)^\mathsf{T}$. Such a transformation is given by

$$\mathtt{H} = \left[\begin{array}{c|c} \mathtt{I} & \mathbf{0} \\ \multicolumn{2}{c}{\boldsymbol{\pi}^\mathsf{T}} \end{array} \right]. \tag{10.2}$$

Indeed, it is immediately verified that $\mathtt{H}^\mathsf{T}(0,0,0,1)^\mathsf{T} = \boldsymbol{\pi}$, and thus $\mathtt{H}^{-\mathsf{T}}\boldsymbol{\pi} = (0,0,0,1)^\mathsf{T}$, as desired. The transformation $\mathtt{H}$ is now applied to all points and the two cameras. Notice, however that this formula will not work if the final coordinate of $\boldsymbol{\pi}^\mathsf{T}$ is zero. In this case, one may compute a suitable $\mathtt{H}$ by computing $\mathtt{H}^{-\mathsf{T}}$ as a Householder matrix (A4.2–*p*580) such that $\mathtt{H}^{-\mathsf{T}}\boldsymbol{\pi} = (0,0,0,1)^\mathsf{T}$.

At this point, the reconstruction that one has is not necessarily the true reconstruction – all one knows is that the plane at infinity is correctly placed. The present reconstruction differs from the true reconstruction by a projective transformation that fixes the plane at infinity. However, according to result 3.7(*p*80), a projective transformation fixing the plane at infinity is an affine transformation. Hence the reconstruction differs by an affine transformation from the true reconstruction – it is an *affine reconstruction.*

An affine reconstruction may well be sufficient for some applications. For example, the mid-point of two points and the centroid of a set of points may now be computed, and lines constructed parallel to other lines and to planes. Such computations are not possible from a projective reconstruction.

As has been stated, the plane at infinity cannot be identified unless some extra information is given. We will now give several examples of the type of information that suffices for this identification.

Translational motion

Consider the case where the camera is known to undergo a purely translational motion. In this case, it is possible to carry out affine reconstruction from two views. A simple way of seeing this is to observe that a point $\mathbf{X}$ on the plane at infinity will map to the same point in two images related by a translation. This is easily verified formally. It is also part of our common experience that as one moves in a straight line (for instance in

Downloaded from https://www.cambridge.org/core. University of Illinois at Urbana - Champaign Library, on 22 Sep 2018 at 15:36:07, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811685.015

a car on a straight road), objects at a great distance (such as the moon) do not appear to move – only the nearby objects move past the field of view. This being so, one may invent any number of matched points $\mathbf{x}_i \leftrightarrow \mathbf{x}_i$ where a point in one image corresponds with the same point in the other image. Note that one does not actually have to observe such a correspondence in the two images – any point and the same point in the other image will do. Given a projective reconstruction, one may then reconstruct the point $\mathbf{X}_i$ corresponding to the match $\mathbf{x}_i \leftrightarrow \mathbf{x}_i$. Point $\mathbf{X}_i$ will lie on the plane at infinity. From three such points one can get three points on the plane at infinity – sufficient to determine it uniquely.

Although this argument gives a constructive proof that affine reconstruction is possible from a translating camera, this does not mean that this is the best way to proceed numerically. In fact in this case, the assumption of translational motion implies a very restricted form for the fundamental matrix – it is skew-symmetric as shown in section 9.3.1. This special form should be taken into account when solving for the fundamental matrix.

Result 10.2. *Suppose the motion of the cameras is a pure translation with no rotation and no change in the internal parameters. As shown in example 9.6(*p249*)* $\mathtt{F} = [\mathbf{e}]_\times = [\mathbf{e}']_\times$, *and for an affine reconstruction one may choose the two cameras as* $\mathtt{P} = [\mathtt{I} \mid \mathbf{0}]$ *and* $\mathtt{P}' = [\mathtt{I} \mid \mathbf{e}']$.

Scene constraints

Scene constraints or conditions may also be used to obtain an affine reconstruction. As long as three points can be identified that are known to lie on the plane at infinity, then that plane may be identified, and the reconstruction transformed to an affine reconstruction.

Parallel lines. The most obvious such condition is the knowledge that 3D lines are in reality parallel. The intersection of the two parallel lines in space gives a point on the plane at infinity. The image of this point is the vanishing point of the line, and is the point of intersection of the two imaged lines. Suppose that three sets of parallel lines can be identified in the scene. Each set intersects in a point on the plane at infinity. Provided each set has a different direction, the three points will be distinct. Since three points determine a plane, this information is sufficient to identify the plane $\boldsymbol{\pi}$.

The best way of actually computing the intersection of lines in space is a somewhat delicate problem, since in the presence of noise, lines that are intended to intersect rarely do. It is discussed in some detail in chapter 12. Correct numerical procedures for computing the plane are given in chapter 13. An example of an affine reconstruction computed from three sets of parallel scene lines is given in figure 10.4.

Note that it is not necessary to find the vanishing point in both images. Suppose the vanishing point $\mathbf{v}$ is computed from imaged parallel lines in the first image, and $\mathbf{l}'$ is a corresponding line in the second image. Vanishing points satisfy the epipolar constraint, so the corresponding vanishing point $\mathbf{v}'$ in the second image may be computed as the intersection of $\mathbf{l}'$ and the epipolar line $\mathtt{F}\mathbf{v}$ of $\mathbf{v}$. The construction of the

Downloaded from https://www.cambridge.org/core. University of Illinois at Urbana - Champaign Library, on 22 Sep 2018 at 15:36:07, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811685.015

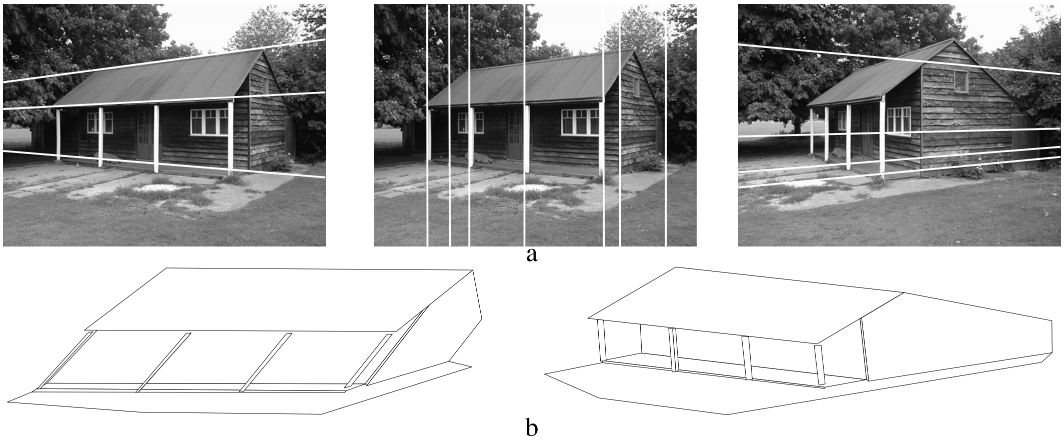

Fig. 10.4. **Affine reconstruction.** *The projective reconstruction of figure 10.3 may be upgraded to affine using parallel scene lines. (a) There are 3 sets of parallel lines in the scene, each set with a different direction. These 3 sets enable the position of the plane at infinity, $\boldsymbol{\pi}_\infty$, to be computed in the projective reconstruction. The wireframe projective reconstruction of figure 10.3 is then affinely rectified using the homography (10.2). (b) Shows two orthographic views of the wireframe affine reconstruction. Note that parallel scene lines are parallel in the reconstruction, but lines that are perpendicular in the scene are not perpendicular in the reconstruction.*

3-space point $\mathbf{X}$ can be neatly expressed algebraically as the solution of the equations $([\mathbf{v}]_\times \mathtt{P})\mathbf{X} = \mathbf{0}$ and $(\mathbf{l}'^\mathsf{T}\mathtt{P}')\mathbf{X} = 0$. These equations expresses the fact that $\mathbf{X}$ maps to $\mathbf{v}$ in the first image, and to a point on $\mathbf{l}'$ in the second image.

Distance ratios on a line. An alternative to computing vanishing points as the intersection of imaged parallel scene lines is to use knowledge of affine length ratios in the scene. For example, given two intervals on a line with a known length ratio, the point at infinity on the line may be determined. This means that from an image of a line on which a world distance ratio is known, for example that three points are equally spaced, the vanishing point may be determined. This computation, and other means of computing vanishing points and vanishing lines, are described in section 2.7(*p*47).

The infinite homography

Once the plane at infinity has been located, so that we have an affine reconstruction, then we also have an image-to-image map called the "infinite homography". This map, which is a 2D homography , is described in greater detail in chapter 13. Briefly, it is the map that transfers points from the $\mathtt{P}$ image to the $\mathtt{P}'$ image via the plane at infinity as follows: the ray corresponding to a point $\mathbf{x}$ is extended to meet the plane at infinity in a point $\mathbf{X}$; this point is projected to a point $\mathbf{x}'$ in the other image. The homography from $\mathbf{x}$ to $\mathbf{x}'$ is written as $\mathbf{x}' = \mathtt{H}_\infty \mathbf{x}$.

Having an affine reconstruction is equivalent to knowing the infinite homography as will now be shown. Given two cameras $\mathtt{P} = [\mathtt{M} \mid \mathbf{m}]$ and $\mathtt{P}' = [\mathtt{M}' \mid \mathbf{m}']$ of an affine reconstruction, the infinite homography is given by $\mathtt{H}_\infty = \mathtt{M}'\mathtt{M}^{-1}$. This is because a point $\mathbf{X} = (\tilde{\mathbf{X}}^\mathsf{T}, 0)^\mathsf{T}$ on the plane at infinity maps to $\mathbf{x} = \mathtt{M}\tilde{\mathbf{X}}$ in one image and $\mathbf{x}' = \mathtt{M}'\tilde{\mathbf{X}}$ in the other, so $\mathbf{x}' = \mathtt{M}'\mathtt{M}^{-1}\mathbf{x}$ for points on $\boldsymbol{\pi}_\infty$. Furthermore, it may be verified that

Downloaded from https://www.cambridge.org/core. University of Illinois at Urbana - Champaign Library, on 22 Sep 2018 at 15:36:07, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811685.015

this is unchanged by a 3-space affine transformation of the cameras. Hence, the infinite homography may be computed explicitly from an affine reconstruction, and vice versa:

Result 10.3. *If an affine reconstruction has been obtained in which the camera matrices are* $\mathtt{P} = [\mathtt{I} \mid \mathbf{0}]$ *and* $\mathtt{P}' = [\mathtt{M}' \mid \mathbf{e}']$*, then the infinite homography is given by* $\mathtt{H}_\infty = \mathtt{M}'$*. Conversely, if the infinite homography* $\mathtt{H}_\infty$ *has been obtained, then the cameras of an affine reconstruction may be chosen as* $\mathtt{P} = [\mathtt{I} \mid \mathbf{0}]$ *and* $\mathtt{P}' = [\mathtt{H}_\infty \mid \mathbf{e}']$.

The infinite homography may be computed directly from corresponding image entities, rather than indirectly from an affine reconstruction. For example, $\mathtt{H}_\infty$ can be computed from the correspondence of three vanishing points together with $\mathtt{F}$, or the correspondence of a vanishing line and vanishing point, together with $\mathtt{F}$. The correct numerical procedure for these computations is given in chapter 13. However, such direct computations are completely equivalent to determining $\boldsymbol{\pi}_\infty$ in a projective reconstruction.

One of the cameras is affine

Another important case in which affine reconstruction is possible is when one of the cameras is known to be an affine camera as defined in section 6.3.1(*p*166). To see that this implies that affine reconstruction is possible, refer to section 6.3.5(*p*172) where it was shown that the principal plane of an affine camera is the plane at infinity. Hence to convert a projective reconstruction to an affine reconstruction, it is sufficient to find the principal plane of the camera supposed to be affine and map it to the plane $(0, 0, 0, 1)^\mathsf{T}$. Recall (section 6.2(*p*158)) that the principal plane of a camera is simply the third row of the camera matrix. For example, consider a projective reconstruction with camera matrices $\mathtt{P} = [\mathtt{I} \mid \mathbf{0}]$ and $\mathtt{P}'$, for which the first camera is supposed to be affine. To map the third row of $\mathtt{P}$ to $(0, 0, 0, 1)$ it is sufficient to swap the last two columns of the two camera matrices, while at the same time swapping the 3rd and 4th coordinates of each $\mathbf{X}_i$. This is a projective transformation corresponding to a permutation matrix $\mathtt{H}$. This shows:

Result 10.4. *Let* $(\mathtt{P}, \mathtt{P}', \{\mathbf{X}_i\})$ *be a projective reconstruction from a set of point correspondences for which* $\mathtt{P} = [\mathtt{I} \mid \mathbf{0}]$*. Suppose in truth,* $\mathtt{P}$ *is known to be an affine camera, then an affine reconstruction is obtained by swapping the last two columns of* $\mathtt{P}$ *and* $\mathtt{P}'$ *and the last two coordinates of each* $\mathbf{X}_i$.

Note that the condition that one of the cameras is affine places no restriction on the fundamental matrix, since any canonical camera pair $\mathtt{P} = [\mathtt{I} \mid \mathbf{0}]$ and $\mathtt{P}'$ can be transformed to a pair in which $\mathtt{P}$ is affine. If both the cameras are known to be affine, then it will be seen that the fundamental matrix has the restricted form given in (14.1–*p*345). In this case, for numerical stability, one must solve for the fundamental matrix enforcing this special form of the fundamental matrix.

Of course there is no such thing as a real affine camera – the affine camera model is an approximation which is only valid when the set of points seen in the image has small depth variation compared with the distance from the camera. Nevertheless, an assumption of an affine camera may be useful to effect the significant restriction from projective to affine reconstruction.

Downloaded from https://www.cambridge.org/core. University of Illinois at Urbana - Champaign Library, on 22 Sep 2018 at 15:36:07, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811685.015

10.4.2 The step to metric reconstruction

Just as the key to affine reconstruction is the identification of the plane at infinity, the key to metric reconstruction is the identification of the absolute conic (section 3.6-(*p*81)). Since the absolute conic, Ω_∞, is a planar conic, lying in the plane at infinity, identifying the absolute conic implies identifying the plane at infinity.

In a stratified approach, one proceeds from projective to affine to metric reconstruction, so one knows the plane at infinity before finding the absolute conic. Suppose one has identified the absolute conic on the plane at infinity. In principle the next step is to apply an affine transformation to the affine reconstruction such that the identified absolute conic is mapped to the absolute conic in the standard Euclidean frame (it will then have the equation $X_1^2 + X_2^2 + X_3^2 = 0$, on $\boldsymbol{\pi}_\infty$). The resulting reconstruction is then related to the true reconstruction by a projective transformation which fixes the absolute conic. It follows from result 3.9(*p*82) that the projective transformation is a similarity transformation, so we have achieved a metric reconstruction.

In practice the easiest way to accomplish this is to consider the image of the absolute conic in one of the images. The image of the absolute conic (as any conic) is a conic in the image. The back-projection of this conic is a cone, which will meet the plane at infinity in a single conic, which therefore defines the absolute conic. Remember that the image of the absolute conic is a property of the image itself, and like any image point, line or other feature, is not dependent on any particular reconstruction, hence it is unchanged by 3D transformations of the reconstruction.

Suppose that in the affine reconstruction the image of the absolute conic as seen by the camera with matrix $\mathtt{P} = [\mathtt{M} \mid \mathbf{m}]$ is a conic $\boldsymbol{\omega}$. We will show how $\boldsymbol{\omega}$ may be used to define the homography $\mathtt{H}$ which transforms the affine reconstruction to a metric reconstruction:

Result 10.5. *Suppose that the image of the absolute conic is known in some image to be $\boldsymbol{\omega}$, and one has an affine reconstruction in which the corresponding camera matrix is given by $\mathtt{P} = [\mathtt{M} \mid \mathtt{m}]$. Then, the affine reconstruction may be transformed to a metric reconstruction by applying a 3D transformation of the form*

$$\mathtt{H} = \begin{bmatrix} \mathtt{A}^{-1} & \\ & 1 \end{bmatrix}$$

where $\mathtt{A}$ is obtained by Cholesky factorization from the equation $\mathtt{A}\mathtt{A}^\mathsf{T} = (\mathtt{M}^\mathsf{T}\boldsymbol{\omega}\mathtt{M})^{-1}$.

Proof. Under the transformation $\mathtt{H}$, the camera matrix $\mathtt{P}$ is transformed to a matrix $\mathtt{P}_\mathrm{M} = \mathtt{P}\mathtt{H}^{-1} = [\mathtt{M}_\mathrm{M} \mid \mathbf{m}_\mathrm{M}]$. If $\mathtt{H}^{-1}$ is of the form

$$\mathtt{H}^{-1} = \begin{bmatrix} \mathtt{A} & \mathbf{0} \\ \mathbf{0}^\mathsf{T} & 1 \end{bmatrix}$$

then $\mathtt{M}_\mathrm{M} = \mathtt{M}\mathtt{A}$. However, the image of the absolute conic is related to the camera matrix $\mathtt{P}_\mathrm{M}$ of a Euclidean frame by the relationship

$$\boldsymbol{\omega}^* = \mathtt{M}_\mathrm{M}\mathtt{M}_\mathrm{M}^\mathsf{T} \ .$$

Downloaded from https://www.cambridge.org/core. University of Illinois at Urbana - Champaign Library, on 22 Sep 2018 at 15:36:07, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811685.015

This is because the camera matrix may be decomposed as $\mathtt{M}_{\mathrm{M}} = \mathtt{KR}$, and from (8.11–*p*210) $\boldsymbol{\omega}^* = \boldsymbol{\omega}^{-1} = \mathtt{KK}^{\mathsf{T}}$. Combining this with $\mathtt{M}_{\mathrm{M}} = \mathtt{MA}$ gives $\boldsymbol{\omega}^{-1} = \mathtt{MAA}^{\mathsf{T}}\mathtt{M}^{\mathsf{T}}$, which may be rearranged as $\mathtt{AA}^{\mathsf{T}} = (\mathtt{M}^{\mathsf{T}}\boldsymbol{\omega}\mathtt{M})^{-1}$. A particular value of $\mathtt{A}$ that satisfies this relationship is found by taking the Cholesky factorization of $(\mathtt{M}^{\mathsf{T}}\boldsymbol{\omega}\mathtt{M})^{-1}$. This latter matrix is guaranteed to be positive-definite (see result A4.5(*p*582)), otherwise no such matrix $\mathtt{A}$ will exist, and metric reconstruction will not be possible. □

This approach to metric reconstruction relies on identifying the image of the absolute conic. There are various ways of doing this and these are discussed next. Three sources of constraint on the image of the absolute conic are given, and in practice a combination of these constraints is used.

1. Constraints arising from scene orthogonality. Pairs of vanishing points, $\mathbf{v}_1$ and $\mathbf{v}_2$, arising from orthogonal scene lines place a single linear constraint on $\boldsymbol{\omega}$:

$$\mathbf{v}_1^{\mathsf{T}}\boldsymbol{\omega}\mathbf{v}_2 = 0.$$

Similarly, a vanishing point $\mathbf{v}$ and a vanishing line $\mathbf{l}$ arising from a direction and plane which are orthogonal place two constraints on $\boldsymbol{\omega}$:

$$\mathbf{l} = \boldsymbol{\omega}\mathbf{v}.$$

A common example is the vanishing point for the vertical direction and a vanishing line from the horizontal ground plane. Finally an imaged scene plane containing metric information, such as a square grid, places two constraints on $\boldsymbol{\omega}$.

2. Constraints arising from known internal parameters. If the calibration matrix of a camera is equal to $\mathtt{K}$, then the image of the absolute conic is $\boldsymbol{\omega} = \mathtt{K}^{-\mathsf{T}}\mathtt{K}^{-1}$. Thus, knowledge of the internal parameters (6.10–*p*157) contained in $\mathtt{K}$ may be used to constrain or determine the elements of $\boldsymbol{\omega}$. In the case where $\mathtt{K}$ is known to have zero skew ($s = 0$),

$$\omega_{12} = \omega_{21} = 0$$

and if the pixels are square (zero skew and $\alpha_x = \alpha_y$) then

$$\omega_{11} = \omega_{22}.$$

These first two sources of constraint are discussed in detail in section 8.8(*p*223) on single view calibration, where examples are given of calibrating a camera solely from such information. Here there is an additional source of constraints available arising from the multiple views.

3. Constraints arising from the same cameras in all images. One of the properties of the absolute conic is that its projection into an image depends only on the calibration matrix of the camera, and not on the position or orientation of the camera. In the case where both cameras $\mathtt{P}$ and $\mathtt{P}'$ have the same calibration matrix (usually meaning that both the images were taken with the same camera with different pose) one has that $\boldsymbol{\omega} = \boldsymbol{\omega}'$, that is the image of the absolute conic is the same in both images. Given

Downloaded from https://www.cambridge.org/core. University of Illinois at Urbana - Champaign Library, on 22 Sep 2018 at 15:36:07, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811685.015

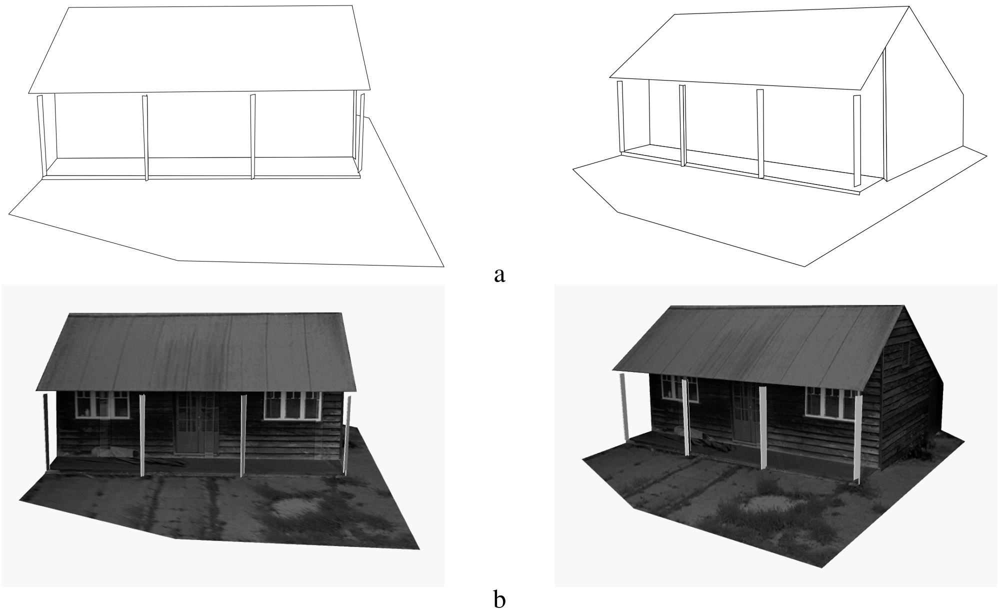

Fig. 10.5. **Metric reconstruction.** *The affine reconstruction of figure 10.4 is upgraded to metric by computing the image of the absolute conic. The information used is the orthogonality of the directions of the parallel line sets shown in figure 10.4, together with the constraint that both images have square pixels. The square pixel constraint is transferred from one image to the other using* $\mathtt{H}_\infty$. *(a) Two views of the metric reconstruction. Lines which are perpendicular in the scene are perpendicular in the reconstruction and also the aspect ratio of the sides of the house is veridical. (b) Two views of a texture mapped piecewise planar model built from the wireframes.*

sufficiently many images, one may use this property to obtain a metric reconstruction from an affine reconstruction. This method of metric reconstruction, and its use for self-calibration of a camera, will be treated in greater detail in chapter 19. For now, we give just the general principle.

Since the absolute conic lies on the plane at infinity, its image may be transferred from one view to the other via the infinite homography. This implies an equation (see result 2.13(*p*37))

$$\boldsymbol{\omega}' = \mathtt{H}_\infty^{-\mathsf{T}} \boldsymbol{\omega} \mathtt{H}_\infty^{-1} \tag{10.3}$$

where $\boldsymbol{\omega}$ and $\boldsymbol{\omega}'$ are images of Ω_∞ in the two views. In forming these equations it is necessary to have an affine reconstruction already, since the infinite homography must be known. If $\boldsymbol{\omega} = \boldsymbol{\omega}'$, then (10.3) gives a set of linear equations in the entries of $\boldsymbol{\omega}$. In general this set of linear equations places four constraints on $\boldsymbol{\omega}$, and since $\boldsymbol{\omega}$ has 5 degrees of freedom it is not completely determined. However, by combining these linear equations with those above provided by scene orthogonality or known internal parameters, $\boldsymbol{\omega}$ may be determined uniquely. Indeed (10.3) may be used to transfer constraints on $\boldsymbol{\omega}$ to constraints on $\boldsymbol{\omega}'$. Figure 10.5 shows an example of a metric reconstruction computed by combining constraints in this manner.

Downloaded from https://www.cambridge.org/core. University of Illinois at Urbana - Champaign Library, on 22 Sep 2018 at 15:36:07, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811685.015

10.4.3 Direct metric reconstruction using ω

The previous discussion showed how knowledge of the image of the absolute conic (IAC) may be used to transform an affine to a metric reconstruction. However, knowing ω it is possible to proceed directly to metric reconstruction, given at least two views. This can be accomplished in at least two different ways. The most evident approach is to use the IAC to compute calibration of each of the cameras, and then carry out a calibrated reconstruction.

This method relies on the connection of ω to the calibration matrix $\mathtt{K}$, namely $\omega = (\mathtt{KK}^\mathsf{T})^{-1}$. Thus one can compute $\mathtt{K}$ from ω by inverting it and then applying Cholesky factorization to obtain $\mathtt{K}$. If the IAC is known in each image, then both cameras may be calibrated in this way. Next with calibrated cameras, a metric reconstruction of the scene may be computed using the essential matrix, as in section 9.6. Note that four possible solutions may result. Two of these are just mirror images, but the other two are different, forming a twisted pair. (Though all solutions but one may be ruled out by consideration of points lying in front of the cameras.)

A more conceptual approach to metric reconstruction is to use knowledge of the IAC to directly determine the plane at infinity and the absolute conic. Knowing the camera matrices $\mathtt{P}$ and $\mathtt{P}'$ in a projective frame, and a conic (specifically the image of the absolute conic) in each image, then Ω_∞ may be explicitly computed in 3-space. This is achieved by back-projecting the conics to cones, which must intersect in the absolute conic. Thus, Ω_∞ and its support plane $\boldsymbol{\pi}_\infty$ are determined (see exercise (x) on page 342 for an algebraic solution). However, two cones will in general intersect in two different plane conics, each lying in a different support plane. Thus there are two possible solutions for the absolute conic, which one can identify as belonging to the two different reconstructions constituting the twisted pair ambiguity.

10.5 Direct reconstruction – using ground truth

It is possible to jump directly from a projective reconstruction to a metric reconstruction if "ground control points" (that is points with known 3D locations in a Euclidean world frame) are given. Suppose we have a set of n such ground control points $\{\mathbf{X}_{\mathrm{E}i}\}$ which are imaged at $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$. We wish to use these points to transform the projective reconstruction to metric.

The 3D location $\{\mathbf{X}_i\}$ of the control points in the projective reconstruction may be computed from their image correspondences $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$. Since the projective reconstruction is related by a homography to the true reconstruction we then have from (10.1) the equations:

$$\mathbf{X}_{\mathrm{E}i} = \mathtt{H}\mathbf{X}_i, \quad i = 1, \ldots, n.$$

Each point correspondence provides 3 linearly independent equations on the elements of $\mathtt{H}$, and since $\mathtt{H}$ has 15 degrees of freedom a linear solution is obtained provided $n \geq 5$ (and no four of the control points are coplanar). This computation, and the proper numerical procedures, are described in chapter 4.

Alternatively, one may bypass the computation of the $\mathbf{X}_i$ and compute $\mathtt{H}$ by relating

Downloaded from https://www.cambridge.org/core. University of Illinois at Urbana - Champaign Library, on 22 Sep 2018 at 15:36:07, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811685.015

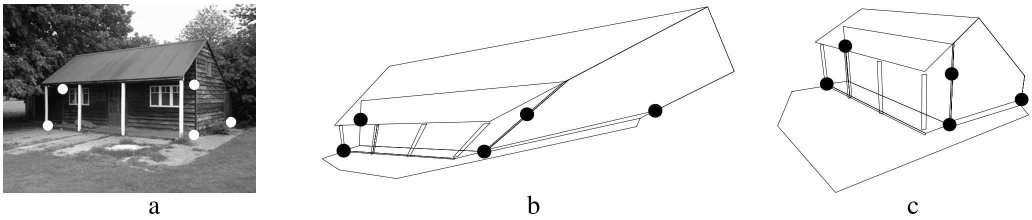

Fig. 10.6. **Direct reconstruction.** *The projective reconstruction of figure 10.3 may be upgraded to metric by specifying the position of five (or more) world points: (a) the five points used; (b) the corresponding points on the projective reconstruction of figure 10.3; (c) the reconstruction after the five points are mapped to their world positions.*

the known ground control points directly to image measurements. Thus as in the DLT algorithm for camera resection (section 7.1(p178)) the equation

$$\mathbf{x}_i = \mathtt{P}\mathtt{H}^{-1}\mathbf{X}_{\mathrm{E}i}$$

provides two linearly independent equations in the entries of the unknown $\mathtt{H}^{-1}$, all other quantities being known. Similarly equations may be derived from the other image if $\mathbf{x}'_i$ is known. It is not necessary for the ground control points to be visible in both images. Note however that if both $\mathbf{x}_i$ and $\mathbf{x}'_i$ *are* visible for a given control point $\mathbf{X}_{\mathrm{E}i}$ then because of the coplanarity constraint on $\mathbf{x}$ and $\mathbf{x}'$, the four equations generated in this way contain only three independent ones.

Once $\mathtt{H}$ has been computed it may be used to transform the cameras $\mathtt{P}, \mathtt{P}'$ of the projective reconstruction to their true Euclidean counterparts. An example of metric structure computed by this direct method is shown in figure 10.6.

10.6 Closure

In this chapter we have overviewed the steps necessary to produce a metric reconstruction from a pair of images. This overview is summarized in algorithm 10.1, and the computational procedures for these steps are described in the following chapters. As usual the general discussion has been restricted mainly to points, but the ideas (triangulation, ambiguity, stratification) apply equally to other image features such as lines, conics etc.

It has been seen that for a metric reconstruction it is necessary to identify two entities in the projective frame; these are the plane at infinity $\boldsymbol{\pi}_\infty$ (for affine), together with the absolute conic Ω_∞ (for metric). Conversely, given $\mathtt{F}$ and a pair of calibrated cameras then $\boldsymbol{\pi}_\infty$ and Ω_∞ may be explicitly computed in 3-space. These entities each have an image-based counterpart: specification of the infinite homography, $\mathtt{H}_\infty$, is equivalent to specifying $\boldsymbol{\pi}_\infty$ in 3-space; and specifying the image of the absolute conic, $\boldsymbol{\omega}$, in each view is equivalent to specifying $\boldsymbol{\pi}_\infty$ and Ω_∞ in 3-space. This equivalence is summarized in table 10.1.

Finally, it is worth noting that if metric precision is not the goal then an acceptable metric reconstruction is generally obtained directly from the projective if approximately correct internal parameters are guessed. Such a "quasi-Euclidean reconstruction" is often suitable for visualization purposes.

Downloaded from https://www.cambridge.org/core. University of Illinois at Urbana - Champaign Library, on 22 Sep 2018 at 15:36:07, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811685.015

Objective

Given two uncalibrated images compute a metric reconstruction $(\mathtt{P}_{\mathrm{M}}, \mathtt{P}'_{\mathrm{M}}, \{\mathbf{X}_{\mathrm{M}i}\})$ of the cameras and scene structure, i.e. a reconstruction that is within a similarity transformation of the true cameras and scene structure.

Algorithm

(i) **Compute a projective reconstruction** $(\mathtt{P}, \mathtt{P}', \{\mathbf{X}_i\})$**:**

- (a) **Compute the fundamental matrix** from point correspondences $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ between the images.
- (b) **Camera retrieval:** compute the camera matrices $\mathtt{P}, \mathtt{P}'$ from the fundamental matrix.
- (c) **Triangulation:** for each point correspondence $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$, compute the point $\mathbf{X}_i$ in space that projects to these two image points.

(ii) **Rectify the projective reconstruction to metric:**

- either **Direct method:** Compute the homography $\mathtt{H}$ such that $\mathbf{X}_{\mathrm{E}i} = \mathtt{H}\mathbf{X}_i$ from five or more ground control points $\mathbf{X}_{\mathrm{E}i}$ with known Euclidean positions. Then the metric reconstruction is

$$\mathtt{P}_{\mathrm{M}} = \mathtt{P}\mathtt{H}^{-1}, \ \ \mathtt{P}'_{\mathrm{M}} = \mathtt{P}'\mathtt{H}^{-1}, \ \ \mathbf{X}_{\mathrm{M}i} = \mathtt{H}\mathbf{X}_i.$$

- or **Stratified method:**
  - (a) **Affine reconstruction:** Compute the plane at infinity, $\boldsymbol{\pi}_\infty$, as described in section 10.4.1, and then upgrade the projective reconstruction to an affine reconstruction with the homography

$$\mathtt{H} = \left[\begin{array}{c} \mathtt{I} \mid \mathbf{0} \\ \boldsymbol{\pi}_\infty^{\mathsf{T}} \end{array} \right].$$

  - (b) **Metric reconstruction:** Compute the image of the absolute conic, $\boldsymbol{\omega}$, as described in section 10.4.2, and then upgrade the affine reconstruction to a metric reconstruction with the homography

$$\mathtt{H} = \left[\begin{array}{cc} \mathtt{A}^{-1} & \\ & 1 \end{array} \right]$$

where $\mathtt{A}$ is obtained by Cholesky factorization from the equation $\mathtt{A}\mathtt{A}^{\mathsf{T}} = (\mathtt{M}^{\mathsf{T}}\boldsymbol{\omega}\mathtt{M})^{-1}$, and $\mathtt{M}$ is the first 3×3 submatrix of the camera in the affine reconstruction for which $\boldsymbol{\omega}$ is computed.

Algorithm 10.1. *Computation of a metric reconstruction from two uncalibrated images.*

10.6.1 The literature

Koenderink and van Doorn [Koenderink-91] give a very elegant discussion of stratification for affine cameras. This was extended to perspective in [Faugeras-95b], with developments given by Luong and Viéville [Luong-94, Luong-96]. The possibility of projective reconstruction given $\mathtt{F}$ appeared in [Faugeras-92b] and Hartley *et al.* [Hartley-92c].

The method of computing affine reconstruction from pure translation first appeared in Moons *et al.* [Moons-94]. Combining scene and internal parameter constraints over multiple views is described in [Faugeras-95c, Liebowitz-99b, Sturm-99c].

Downloaded from https://www.cambridge.org/core. University of Illinois at Urbana - Champaign Library, on 22 Sep 2018 at 15:36:07, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811685.015

Image information provided	View relations and projective objects	3-space objects	Reconstruction ambiguity
Point correspondences	$\mathtt{F}$		Projective
Point correspondences including vanishing points	$\mathtt{F}, \mathtt{H}_\infty$	$\boldsymbol{\pi}_\infty$	Affine
Point correspondences and internal camera calibration	$\mathtt{F}, \mathtt{H}_\infty$ $\boldsymbol{\omega}, \boldsymbol{\omega}'$	$\boldsymbol{\pi}_\infty$ Ω_∞	Metric

Table 10.1. *The two-view relations, image entities, and their 3-space counterparts for various classes of reconstruction ambiguity.*

10.6.2 Notes and exercises

(i) Using only (implicit) image relations (i.e. without an explicit 3D reconstruction) and given the images of a line $\mathbf{L}$ and point $\mathbf{X}$ (not on $\mathbf{L}$) in two views, together with $\mathtt{H}_\infty$ between the views, compute the image of the line in 3-space parallel to $\mathbf{L}$ and through $\mathbf{X}$. Other examples of this implicit approach to computation are given in [Zeller-96].

Downloaded from https://www.cambridge.org/core. University of Illinois at Urbana - Champaign Library, on 22 Sep 2018 at 15:36:07, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811685.015