Single View Metrology

A. CRIMINISI, I. REID AND A. ZISSERMAN *

[criminis,ian,az]@robots.oxford.ac.uk

Department of Engineering Science, University of Oxford, Parks Road, Oxford OX1 3PJ, UK

Abstract. *We describe how 3D affine measurements may be computed from a single perspective view of a scene given only minimal geometric information determined from the image. This minimal information is typically the vanishing line of a reference plane, and a vanishing point for a direction not parallel to the plane. It is shown that affine scene structure may then be determined from the image, without knowledge of the camera's internal calibration (e.g. focal length), nor of the explicit relation between camera and world (pose).*

In particular, we show how to (i) compute the distance between planes parallel to the reference plane (up to a common scale factor); (ii) compute area and length ratios on any plane parallel to the reference plane; (iii) determine the camera's location. Simple geometric derivations are given for these results. We also develop an algebraic representation which unifies the three types of measurement and, amongst other advantages, permits a first order error propagation analysis to be performed, associating an uncertainty with each measurement.

We demonstrate the technique for a variety of applications, including height measurements in forensic images and 3D graphical modelling from single images.

Keywords: 3D reconstruction, video metrology, photogrammetry.

1. Introduction

In this paper we describe how aspects of the affine 3D geometry of a scene may be measured from a single perspective image. We will concentrate on scenes containing planes and parallel lines, although the methods are not so restricted. The methods we develop extend and generalize previous results on single view metrology [RZ96, HAA97, KSH98, PTVG98].

It is assumed that images are obtained by perspective projection. In addition, we assume that the vanishing line of a *reference plane* in the scene may be determined from the image, together with a vanishing point for another *reference direction* (not parallel to the plane). We are then concerned with three canonical types of measurement: (i) measurements of the distance *between* any of the planes which are parallel to the reference plane; (ii) measurements *on* these planes (and comparison of these measurements to those obtained on any parallel plane); and (iii) determining the camera's position in terms of the reference plane and direction.

The measurement methods developed here are independent of the camera's internal parameters:

focal length, aspect ratio, principal point, skew. The camera is always assumed to be uncalibrated, its internal parameters unknown. We analyse situations where the camera (the projection matrix) can only be partially determined from scene landmarks. This is an intermediate situation between calibrated reconstruction (where metric entities like angles between rays can be computed) and completely uncalibrated cameras (where a reconstruction can be obtained only up to a projective transformation).

The ideas in this paper can be seen as reversing the rules for drawing perspective images given by Leon Battista Alberti [Alb35] in his treatise on perspective (1435). These are the rules followed by the Italian Renaissance painters of the 15th century, and indeed we demonstrate the correctness of their mastery of perspective by analysing a painting by Piero della Francesca.

This paper extends the work in [CRZ99b]. Here particular attention is paid to: computing Maximum Likelihood estimates of measurements when more than the minimum number of references are available; transferring measurements from one reference plane to another by making use of planar homologies; analysing in detail the uncertainty of the computed distances; validating the analytical uncertainty predictions by using statistical tests. A number of worked examples are presented to explain the algorithms step by step and demonstrate their validity.

We begin in section 2 by giving simple geometric derivations of how, in principle, three dimensional affine information may be extracted from the image (fig. 1). In section 3 we introduce an algebraic representation of the problem and show that this representation unifies the three canonical measurement types, leading to simple formulae in each case. In section 4 we describe how errors in image measurements propagate to errors in the 3D measurements, and hence we are able to compute confidence intervals on the 3D measurements, i.e. a quantitative assessment of accuracy. The work has a variety of applications, and we demonstrate three important ones: forensic measurement, virtual modelling and furniture measurements in section 5.
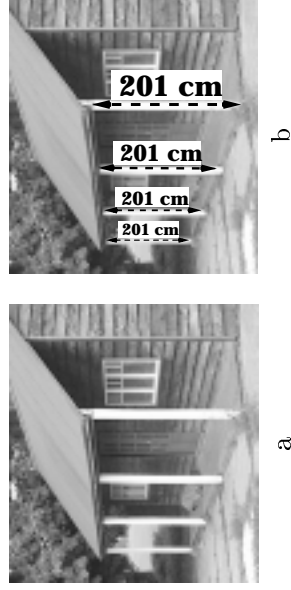

Fig. 1. **Measuring distances of points from a reference plane (the ground) in a single image:** (a) The four pillars have the same height in the world, although their images clearly are not of the same length due to perspective effects. (b) As shown, however, all pillars are correctly measured to have the same height.

2. Geometry

The camera model employed here is central projection. We assume that the vanishing line of a reference plane in the scene may be computed from image measurements, together with a vanishing point for another direction (not parallel to the plane). This information is generally easily obtainable from images of structured scenes [CW90, MK95, LZ98, Shu99]. Effects such as radial distortion (often arising in slightly wide-angle lenses typically used in security cameras) which corrupt the central projection model can generally be removed [DF95], and are therefore not detrimental to our methods. Implementation details for computation of vanishing points and lines, and line detection are given in Appendix A.

Although the schematic figures show the camera centre at a finite location, the results we derive apply also to the case of a camera centre at infinity, i.e. where the images are obtained by parallel projection.

The basic geometry of the plane's vanishing line and the vanishing point are illustrated in figure 2. The vanishing line **l** of the reference plane is the projection of the line at infinity of the reference plane into the image. The vanishing point **v** is the image of the point at infinity in the reference direction. Note that the reference direction need not be vertical, although for clarity we will often refer to the vanishing point as the "vertical" vanishing point. The vanishing point is then the image of the vertical "footprint" of the camera centre on the reference plane. Likewise, the reference plane will often, but not necessarily, be the
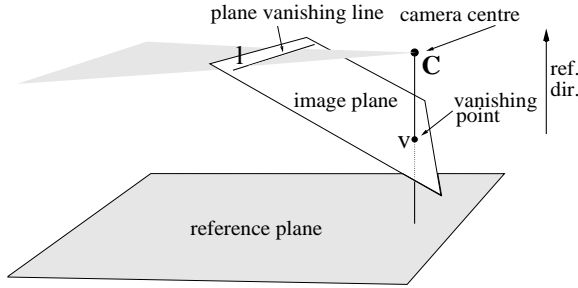
Fig. 2. **Basic geometry**: The plane's vanishing line **l** is the intersection of the image plane with a plane parallel to the reference plane and passing through the camera centre **C**. The vanishing point **v** is the intersection of the image plane with a line parallel to the reference direction through the camera centre.

ground plane, in which case the vanishing line is more commonly known as the "horizon".

It can be seen (for example, by inspection of figure 2) that the vanishing line partitions all points in scene space. Any scene point which projects onto the vanishing line is at the same distance from the plane as the camera centre; if it lies "above" the line it is farther from the plane, and if "below" the vanishing line, then it is closer to the plane than the camera centre.

2.1. *Measurements between parallel planes*

We wish to measure the distance (in the reference direction) between two parallel planes, specified by the image points $\mathbf{x}$ and $\mathbf{x}'$. Figure 3 shows the geometry, with points $\mathbf{x}$ and $\mathbf{x}'$ in correspondence. We use upper case letters ($\mathbf{X}$) to indicate quantities in space and lower case letters ($\mathbf{x}$) to indicate image quantities.

Definition 1. Two points $\mathbf{X}, \mathbf{X}'$ on separate planes (parallel to the reference plane) *correspond* if the line joining them is parallel to the reference direction.

Hence the images of corresponding points and the vanishing point are collinear. For example, if the direction is vertical, then the top of an upright person's head and the sole of his/her foot correspond. If the world distance between the two points is known, we term this a reference distance.
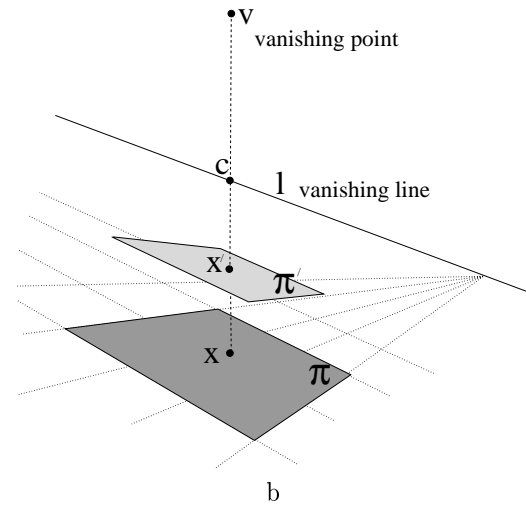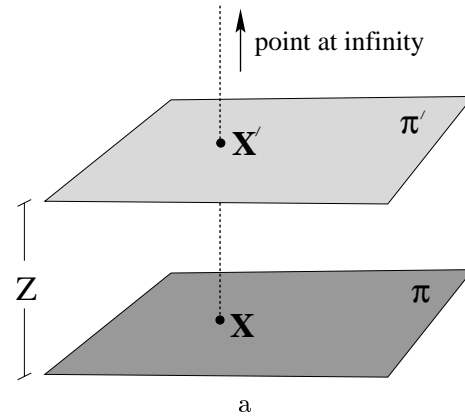
We show that:

Fig. 3. **Distance between two planes relative to the distance of the camera centre from one of the two planes**: (a) in the world; (b) in the image. The point $\mathbf{x}$ on the plane $\boldsymbol{\pi}$ corresponds to the point $\mathbf{x}'$ on the plane $\boldsymbol{\pi}'$. The four aligned points $\mathbf{v}$, $\mathbf{x}$, $\mathbf{x}'$ and the intersection $\mathbf{c}$ of the line joining them with the vanishing line define a cross-ratio. The value of the cross-ratio determines a ratio of distances between planes in the world, see text.

Theorem 1. *Given the vanishing line of a reference plane and the vanishing point for a reference direction, then distances from the reference plane parallel to the reference direction can be computed from their imaged end points up to a common scale factor. The scale factor can be determined from one known reference length.*

Proof: The four points $\mathbf{x}$, $\mathbf{x}'$, $\mathbf{c}$, $\mathbf{v}$ marked on figure 3b define a cross-ratio [Spr64]. The vanishing point is the image of a point at infinity in the scene and the point $\mathbf{c}$, since it lies on the vanishing
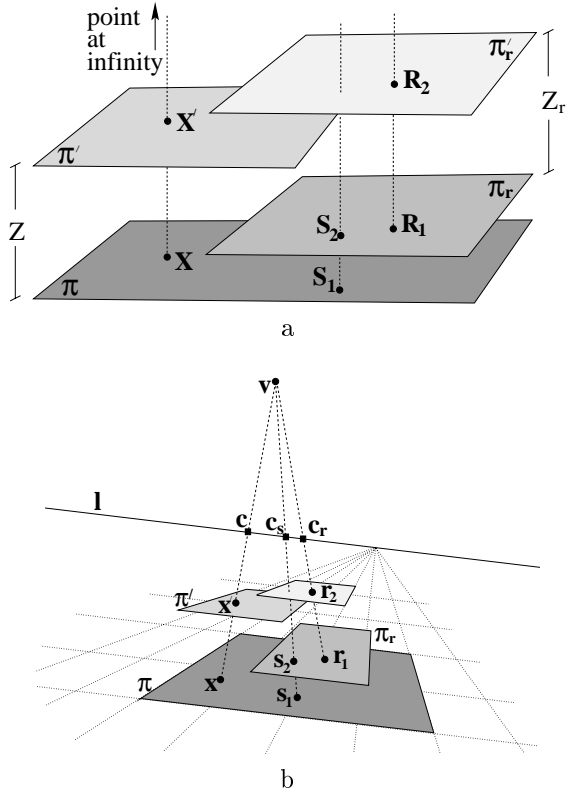
a

b

Fig. 4. **Distance between two planes relative to the distance between two other planes**: (a) in the world; (b) in the image. The point $\mathbf{x}$ on the plane $\boldsymbol{\pi}$ corresponds to the point $\mathbf{x}'$ on the plane $\boldsymbol{\pi}'$. The point $\mathbf{s}_1$ corresponds to the point $\mathbf{s}_2$. The point $\mathbf{r}_1$ corresponds to the point $\mathbf{r}_2$. The distance Z_r in the world between $\mathbf{R}_1$ and $\mathbf{R}_2$ is known and used as reference to compute the distance Z, see text.

line, is the image of a point at distance Z_c from the plane $\boldsymbol{\pi}$, where Z_c is the distance of the camera centre from $\boldsymbol{\pi}$. In the world the value of the cross-ratio provides an affine length ratio which determines the distance Z between the planes containing $\mathbf{X}'$ and $\mathbf{X}$ (in fig. 3a) relative to the camera's distance Z_c from the plane $\boldsymbol{\pi}$ (or $\boldsymbol{\pi}'$ depending on the ordering of the cross-ratio). Note that the distance Z can alternatively be computed using a line-to-line homography avoiding the ordering ambiguity of the cross-ratio.

For the case in figure 3b we can write

$$\frac{d(\mathbf{x}, \mathbf{c})\, d(\mathbf{x}', \mathbf{v})}{d(\mathbf{x}', \mathbf{c})\, d(\mathbf{x}, \mathbf{v})} = \frac{d(\mathbf{X}, \mathbf{C})\, d(\mathbf{X}', \mathbf{V})}{d(\mathbf{X}', \mathbf{C})\, d(\mathbf{X}, \mathbf{V})} \qquad (1)$$

where $d(\mathbf{x}_1, \mathbf{x}_2)$ is distance between two generic points $\mathbf{x}_1$ and $\mathbf{x}_2$. Since the back projection of the point $\mathbf{v}$ is a point at infinity $\frac{d(\mathbf{X}', \mathbf{V})}{d(\mathbf{X}, \mathbf{V})} = 1$ and there-

fore the right hand side of (1) reduces to $\frac{Z_c}{Z_c - Z}$. Simple algebraic manipulation on (1) yields

$$\frac{Z}{Z_c} = 1 - \frac{d(\mathbf{x}', \mathbf{c})\, d(\mathbf{x}, \mathbf{v})}{d(\mathbf{x}, \mathbf{c})\, d(\mathbf{x}', \mathbf{v})} \qquad (2)$$

The absolute distance Z can be obtained from this distance ratio once the camera's distance Z_c is specified.

However it is usually more practical to determine the distance Z via a second measurement in the image, that of a known reference length. In fact, given a known reference distance Z_r, from (2) we can compute the distance of the camera Z_c and then apply (2) to a new pair of end points and compute the distance Z. $\square$

We now generalize Theorem 1 to the following.

Definition 2. A set of parallel planes are *linked* if it is possible to go from one plane to any other plane in the set through a chain of pairs of *corresponding* points (see also definition 1).

For example in fig. 4a the planes $\boldsymbol{\pi}'$, $\boldsymbol{\pi}$, $\boldsymbol{\pi}_r$ and $\boldsymbol{\pi}'_r$ are linked by the chain of correspondences $\mathbf{X}' \leftrightarrow \mathbf{X}$, $\mathbf{S}_1 \leftrightarrow \mathbf{S}_2$, $\mathbf{R}_1 \leftrightarrow \mathbf{R}_2$.

Theorem 2. *Given a set of linked parallel planes, the distance between* any *pair of planes is sufficient to determine the absolute distance between any other pair, the link being provided by a chain of point correspondences between the set of planes.*

Proof: Figure 4 shows a diagram where four parallel planes are imaged. Note that they all share the same vanishing line which is the image of the axis of the pencil. The distance Z_r between two of them can be used as reference to compute the distance Z between the other two as follows:

- From the cross-ratio defined by the four aligned points $\mathbf{v}$, $\mathbf{c}_r$, $\mathbf{r}_2$, $\mathbf{r}_1$ and the known distance Z_r between the points $\mathbf{R}_1$ and $\mathbf{R}_2$ we can compute the distance of the camera from the plane $\boldsymbol{\pi}_r$.
- That camera distance and the cross-ratio defined by the four aligned points $\mathbf{v}$, $\mathbf{c}_s$, $\mathbf{s}_2$, $\mathbf{s}_1$, determine the distance between the planes $\boldsymbol{\pi}_r$ and $\boldsymbol{\pi}$. The distance Z_c of the camera from the plane $\boldsymbol{\pi}$ is, therefore, determined too.
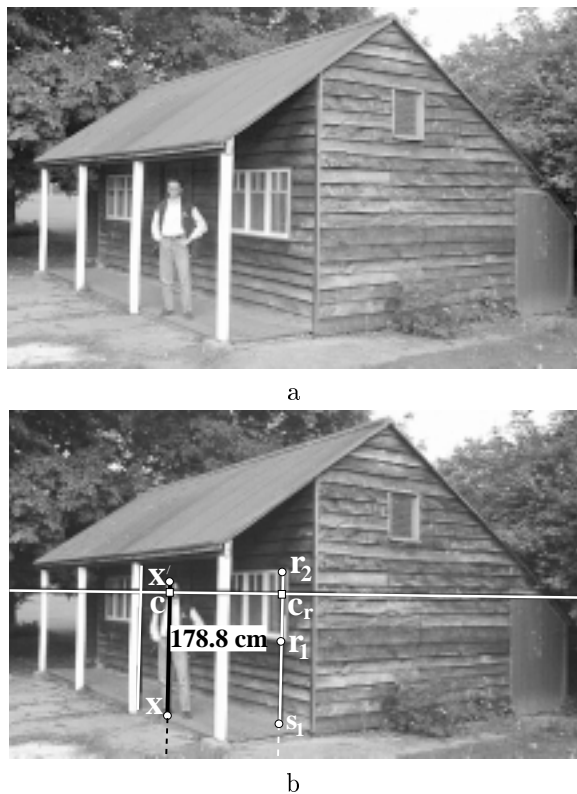
a

b

Fig. 5. **Measuring the height of a person from single view**: (a) original image; (b) the height of the person is computed from the image as 178.8cm; the true height is 180cm, but note that the person is leaning down a bit on his right foot. The vanishing line is shown in white; the vertical vanishing point is not shown since it lies well below the image. The reference distance is in white (the height of the window frame on the right). Compare the marked points with the ones in fig. 4.

- The distance Z_c can now be used in (2) to compute the distance Z between the two planes π and π'.

$\square$

In section 3.1 we give an algebraic derivation of these results which avoids the need to compute the distance of the camera explicitly and simplifies the measurement procedure.

Example. Figure 5 shows that a person's height may be computed from an image given a vertical reference distance elsewhere in the scene. The ground plane is reference. The height of the frame of the window has been measured on site and used as the reference distance (it corresponds to the distance between $\mathbf{R}_1$ and $\mathbf{R}_2$ in the world in fig. 4a). This situation corresponds to the one in figure 4 where the two points $\mathbf{S}_2$ and $\mathbf{R}_1$ (and therefore $\mathbf{s}_2$ and $\mathbf{r}_1$) coincide. The height of the person is computed from the cross ratio defined by the points $\mathbf{x}'$, $\mathbf{c}$, $\mathbf{x}$ and the vanishing point (*c.f.* figure 4b) as described in the proof above. Since the points $\mathbf{S}_2$ and $\mathbf{R}_1$ coincide the derivation is simpler.

2.2. Measurements on parallel planes

If the reference plane π is affine calibrated (we know its vanishing line) then from image measurements we can compute:

1. ratios of lengths of parallel line segments on the plane;
2. ratios of areas on the plane.

Moreover the vanishing line is shared by the pencil of planes parallel to the reference plane, hence affine measurements may be obtained for any other plane in the pencil. However, although affine measurements, such as an area ratio, may be made *on* a particular plane, the areas of regions lying on two parallel planes cannot be compared directly. If the region is parallel projected in the scene from one plane onto the other, affine measurements can then be made from the image since both regions are now on the same plane, and parallel projection between parallel planes does not alter affine properties.

A map in the world between parallel planes induces a projective map in the image between images of points on the two planes. This image map is a *planar homology* [Spr64], which is a plane projective transformation with five degrees of freedom, having a line of fixed points called the *axis*, and a distinct fixed point not on the axis known as the *vertex*. Planar homologies arise naturally in an image when two planes related by a perspectivity in three-dimensional space are imaged [VGPZ98]. The geometry is illustrated in figure 7.

In our case the vanishing line of the plane, and the vertical vanishing point, are, respectively, the axis and vertex of the homology which relates a pair of planes in the pencil.
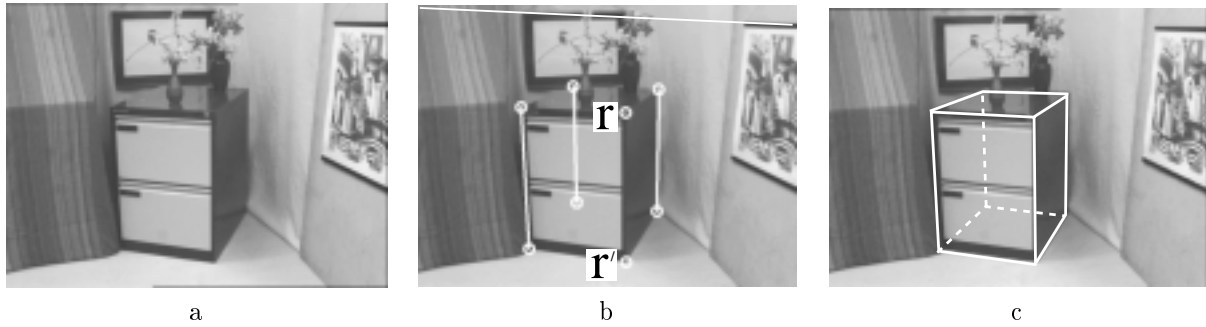
a   b   c

Fig. 6. **Homology mapping of points from one plane to a parallel one**: (a) original image, the floor and the top of the filing cabinet are parallel planes. (b) Their common vanishing line (axis of the homology, shown in white) has been computed by intersecting two sets of horizontal edges. The vertical vanishing point (vertex of the homology) has been computed by intersecting vertical edges. Two corresponding points **r** and **r**′ are selected and the homology computed. Three corners of the top plane of the cabinet have been selected and their corresponding points on the floor computed by the homology. Note that occluded corners have been retrieved too. (c) The wire frame model shows the structure of the cabinet; occluded sides are dashed.

The homology can then be parametrized as [VL99]

$$\tilde{\mathtt{H}} = \mathtt{I} + \mu \frac{\mathbf{v}\mathbf{l}^\top}{\mathbf{v}\cdot\mathbf{l}} \qquad (3)$$

where **v** is the vanishing point, **l** is the plane vanishing line and μ is a scale factor. Thus **v** and **l** specify four of the five degrees of freedom of the homology. The remaining degree of freedom of the homology, μ, is uniquely determined from any pair of image points which correspond between the planes (points **r** and **r**′ in figure 7).

Once the matrix $\tilde{\mathtt{H}}$ is computed each point on a plane can be transferred into the corresponding point on a parallel plane as $\mathbf{x}' = \tilde{\mathtt{H}}\mathbf{x}$. An example of this *homology mapping* is shown in figure 6.

Consequently we can compare measurements made on two separate planes. In particular we may compute:

1. the ratio between two lengths measured along parallel lines, one length on each plane;
2. the ratio between two areas, one area on each plane.

In fact we can simply transfer all points from one plane to the reference plane using the homology and then, since the reference plane's vanishing line is known we may make affine measurements in the plane, e.g. ratios of lengths on parallel lines or ratios of areas.

Example. Figure 8 shows an example. The vanishing line of the two front facing walls and the vanishing point are known as is the point corre-

spondence **r**, **r**′ in the reference direction. The ratio of lengths of parallel line segments is computed by using formulae given in section 3.2.

Notice that errors in the selection of point positions affect the computations; the veridical values of the ratios in figure 8 are exact integers. A proper error analysis is necessary to estimate the uncertainty of these affine measurements.

2.3. Determining the camera position

In section 2.1, we computed distances between planes as a ratio relative to the camera's distance from the reference plane. Conversely, we may compute the camera's distance Z_c from a particular plane knowing a single reference distance Z_r.

Furthermore, by considering figure 2 it is seen that the location of the camera relative to the reference plane is the back-projection of the vertical vanishing point onto the reference plane. This back-projection is accomplished by a homography which maps the image to the reference plane (and vice-versa). Although the choice of coordinate frame in the world is somewhat arbitrary, fixing this frame immediately defines the homography uniquely and hence the camera position.

3. Algebraic Representation

The measurements described in the previous section are computed in terms of cross-ratios. In this section we develop a uniform algebraic approach
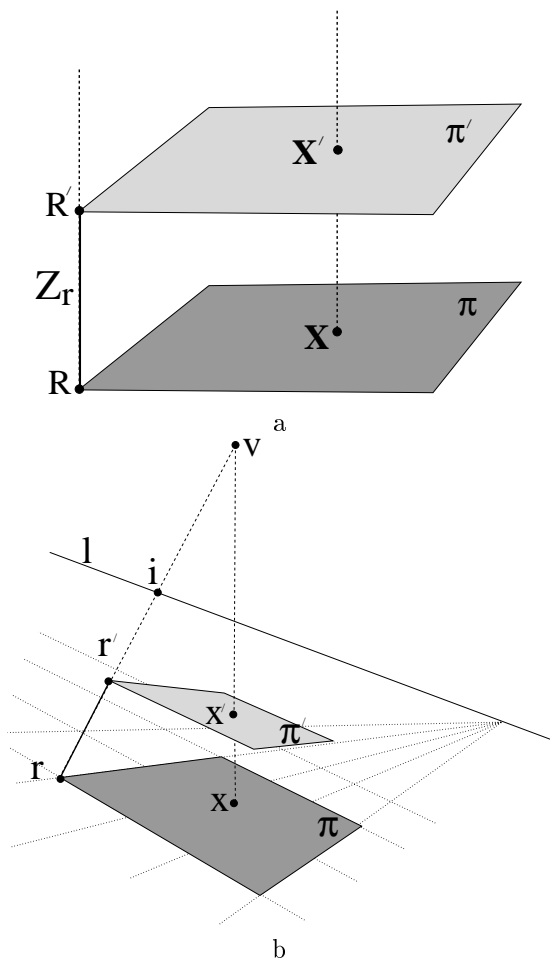
Fig. 7. **Homology mapping between imaged parallel planes**: (a) A point $\mathbf{X}$ on plane π is mapped into the point $\mathbf{X}'$ on π' by a parallel projection. (b) In the image the mapping between the images of the two planes is a homology, where $\mathbf{v}$ is the *vertex* and $\mathbf{l}$ the *axis*. The correspondence $\mathbf{r} \rightarrow \mathbf{r}'$ fixes the remaining degree of freedom of the homology from the cross-ratio of the four points: $\mathbf{v}$, $\mathbf{i}$, $\mathbf{r}'$ and $\mathbf{r}$.

to the problem which has a number of advantages over direct geometric construction: first, it avoids potential problems with ordering for the cross-ratio; second, it enables us to deal with both minimal or over-constrained configurations uniformly; third, we unify the different types of measurement within one representation; and fourth, in section 4 we use this algebraic representation to develop an uncertainty analysis for measurements.

To begin we define an affine coordinate system XYZ in space [KvD91, QM92]. Let the origin of the coordinate frame lie on the reference plane,

Fig. 8. **Measuring ratio of lengths of parallel line segments lying on two parallel scene planes:** The points $\mathbf{r}$ and $\mathbf{r}'$ (together with the plane vanishing line and the vanishing point) define the homology between the two planes on the facade of the building.

with the X and Y-axes spanning the plane. The Z-axis is the reference direction, which is thus any direction not parallel to the plane. The image coordinate system is the usual xy affine image frame, and a point $\mathbf{X}$ in space is projected to the image point $\mathbf{x}$ via a 3×4 projection matrix P as:

$$\mathbf{x} = \mathsf{P}\mathbf{X} = \begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_3 & \mathbf{p}_4 \end{bmatrix} \mathbf{X}$$

where $\mathbf{x}$ and $\mathbf{X}$ are homogeneous vectors in the form: $\mathbf{x} = (x, y, w)^\top$, $\mathbf{X} = (X, Y, Z, W)^\top$, and "$=$" means equality up to scale.

If we denote the vanishing points for the X, Y and Z directions as (respectively) $\mathbf{v}_X$, $\mathbf{v}_Y$ and $\mathbf{v}$, then it is clear by inspection [Fau93] that the first three columns of P are the vanishing points: $\mathbf{v}_X = \mathbf{p}_1$, $\mathbf{v}_Y = \mathbf{p}_2$ and $\mathbf{v} = \mathbf{p}_3$, and that the final column of P is the projection of the origin of the world coordinate system, $\mathbf{o} = \mathbf{p}_4$. Since our choice of coordinate frame has the X and Y axes in the reference plane $\mathbf{p}_1 = \mathbf{v}_X$ and $\mathbf{p}_2 = \mathbf{v}_Y$ are two distinct points on the vanishing line. Choosing these fixes the X and Y affine coordinate axes. We denote the vanishing line by $\mathbf{l}$, and to emphasize that the vanishing points $\mathbf{v}_X$ and $\mathbf{v}_Y$ lie on it, we denote them by $\mathbf{l}_1^\perp$, $\mathbf{l}_2^\perp$, with $\mathbf{l}_i^\perp \cdot \mathbf{l} = 0$.

Columns 1, 2 and 4 of the projection matrix are the three columns of the reference plane to image homography. This homography must have rank
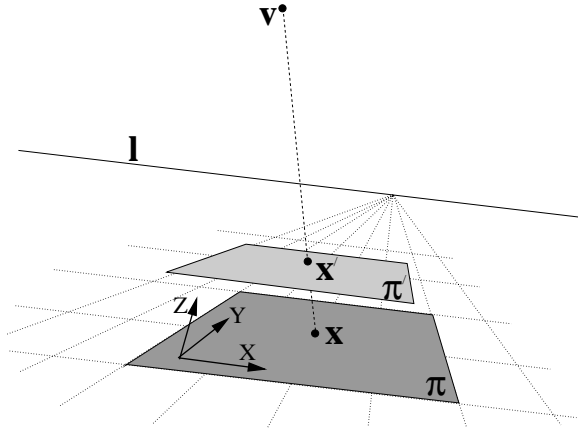
Fig. 9. Measuring the distance of a plane $\boldsymbol{\pi}'$ from the parallel reference plane $\boldsymbol{\pi}$, the geometry.

three, otherwise the reference plane to image map is degenerate. Consequently, the final column (the origin of the coordinate system) must not lie on the vanishing line, since if it does then all three columns are points on the vanishing line, and thus are not linearly independent. Hence we set it to be $\mathbf{p}_4 = \mathbf{l}/\|\mathbf{l}\| = \bar{\mathbf{l}}$.

Therefore the final parameterization of the projection matrix P is:

$$P = \begin{bmatrix} \mathbf{l}_1^\perp & \mathbf{l}_2^\perp & \alpha\mathbf{v} & \bar{\mathbf{l}} \end{bmatrix} \qquad (4)$$

where α is a scale factor, which has an important rôle to play in the remainder of the paper.

Note that the vertical vanishing point $\mathbf{v}$ imposes two constraints on the P matrix, the vanishing line $\mathbf{l}$ imposes two and the α parameter only one for a total of five independent constraints (at this stage the first two columns of the P matrix are not completely known; the only constraint is that they are orthogonal to the plane vanishing line $\mathbf{l}$, $\mathbf{l}_i^\top \cdot \mathbf{l} = 0$). In general however the P matrix has eleven d.o.f., which can be regarded as comprising eight for the world-to-image homography induced by the reference plane, two for the vanishing point and one for the affine parameter α. In our case the vanishing line determines two of the eight d.o.f. of the homography.

In the following sections we show how to compute various measurements from this projection matrix. Measurements of distances between planes are independent of the first two (in general under-determined) columns of P. If $\mathbf{v}$ and $\mathbf{l}$ are specified, the only unknown quantity for these measurements is α. Coordinate measure-

ments within the planes depend on the first two and the fourth columns of P. These columns define an affine coordinate frame within the plane. Affine measurements (e.g. area ratios), though, are independent of the actual coordinate frame and depend only on the fourth column of P. If any metric information on the plane is known, we may impose constraints on the choice of the frame.

3.1. Measurements between parallel planes

3.1.1. Distance of a plane from the reference plane $\boldsymbol{\pi}$

We wish to measure the distance between scene planes specified by a point $\mathbf{X}$ and a point $\mathbf{X}'$ in the scene (see fig. 3a). These points may be chosen as respectively $\mathbf{X} = (X, Y, 0)^\top$ and $\mathbf{X}' = (X, Y, Z)^\top$, and their images are $\mathbf{x}$ and $\mathbf{x}'$ (fig. 9). If P is the projection matrix then the image coordinates are

$$\mathbf{x} = P \begin{pmatrix} X \\ Y \\ 0 \\ 1 \end{pmatrix}, \qquad \mathbf{x}' = P \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

The equations above can be rewritten as

$$\mathbf{x} = \rho(X\mathbf{p}_1 + Y\mathbf{p}_2 + \mathbf{p}_4) \qquad (5)$$
$$\mathbf{x}' = \rho'(X\mathbf{p}_1 + Y\mathbf{p}_2 + Z\mathbf{p}_3 + \mathbf{p}_4) \qquad (6)$$

where ρ and ρ' are unknown scale factors, and $\mathbf{p}_i$ is the ith column of the P matrix.

Since $\mathbf{p}_1 \cdot \bar{\mathbf{l}} = \mathbf{p}_2 \cdot \bar{\mathbf{l}} = 0$ and $\mathbf{p}_4 \cdot \bar{\mathbf{l}} = 1$, taking the scalar product of (5) with $\bar{\mathbf{l}}$ yields $\rho = \bar{\mathbf{l}} \cdot \mathbf{x}$ and therefore (6) can be rewritten as

$$\mathbf{x}' = \rho'(\frac{\mathbf{x}}{\rho} + \alpha Z\mathbf{v}) \qquad (7)$$

By taking the vector product of both terms of (7) with $\mathbf{x}'$ we obtain

$$\mathbf{x} \times \mathbf{x}' = -\alpha Z\rho(\mathbf{v} \times \mathbf{x}')) \qquad (8)$$

and, finally, taking the norm of both sides of (8) yields

$$\alpha Z = -\frac{\|\mathbf{x} \times \mathbf{x}'\|}{(\bar{\mathbf{l}} \cdot \mathbf{x})\|\mathbf{v} \times \mathbf{x}'\|} \qquad (9)$$

Since αZ scales linearly with α, affine structure has been obtained. If α is known, then a metric value for Z can be immediately computed as:

$$Z = -\frac{\|\mathbf{x} \times \mathbf{x}'\|}{(\mathbf{p}_4 \cdot \mathbf{x})\|\mathbf{p}_3 \times \mathbf{x}'\|} \qquad (10)$$

Conversely, if Z is known (i.e. it is a reference distance) then (9) provides a means of computing α, and hence removing the affine ambiguity.

Metric calibration from multiple references.

If more than one reference distance is known then an estimate of α can be derived from an error minimization algorithm. We here show a special case where all distances are measured from the same reference plane and an algebraic error is minimized. An optimal minimization algorithm will be described in section 4.2.1.

For the ith reference distance Z_i with end points $\mathbf{r}_i$ and $\mathbf{r}'_i$ we define: $\beta_i = \|\mathbf{r}_i \times \mathbf{r}'_i\|$, $\rho_i = \bar{\mathbf{l}} \cdot \mathbf{r}_i$, $\gamma_i = \|\mathbf{v} \times \mathbf{r}'_i\|$. Therefore, from (9) we obtain:

$$\alpha Z \rho_i \gamma_i = -\beta_i \qquad (11)$$

Note that all the points $\mathbf{r}_i$ are images of world points $\mathbf{R}_i$ on the reference plane π.

We now define the $n \times 2$ matrix $\mathtt{A}$ (reorganising (11)) as:

$$\mathtt{A} = \begin{pmatrix} Z_1 \rho_1 \gamma_1 & \beta_1 \\ \vdots & \vdots \\ Z_i \rho_i \gamma_i & \beta_i \\ \vdots & \vdots \\ Z_n \rho_n \gamma_n & \beta_n \end{pmatrix}$$

where n is the number of reference distances.

If there is no measurement error or $n = 1$ then $\mathtt{A}\mathbf{s} = \mathbf{0}$ where $\mathbf{s} = (\, s_1 \;\; s_2 \,)^\top$ is a homogeneous 2-vector and

$$\alpha = \frac{s_1}{s_2} \qquad (12)$$

In general $n > 1$ and uncertainty is present in the reference distances. In this case we find the solution $\mathbf{s}$ which minimizes $\|\mathtt{A}\mathbf{s}\|$. That is the eigenvector of the 2×2 matrix $\mathtt{M} = \mathtt{A}^\top \mathtt{A}$ corresponding to its minimum eigenvalue. The parameter α is finally computed from (12).

With more reference distances Z_i, α is estimated more accurately (see section 4), but no more constraints are added on the $\mathtt{P}$ matrix.
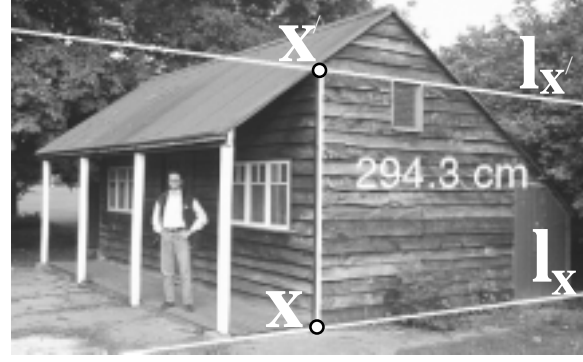

Fig. 10. **Measuring heights using parallel lines:** The vertical vanishing point and the vanishing line for the ground plane have been computed. The distance of the top of the window on the left wall from the ground is known and used as reference. The distance of the top of the window on the right wall from the ground is computed from the distance between the two horizontal lines whose images are $\mathbf{l}_{x'}$ and $\mathbf{l}_x$. The top line $\mathbf{l}_{x'}$ is defined by the top edge of the window, and the line $\mathbf{l}_x$ is the corresponding one on the ground plane. The distance between them is computed to be 294.3cm.

Worked example. In figure 10 the distance of a horizontal line from the ground is measured.

- The vertical vanishing point $\mathbf{v}$ is computed by intersecting vertical (scene) edges;

All images of lines parallel to the ground plane intersect in points on the horizon, therefore:

- A point $\mathbf{v}_1$ on the horizon is computed by intersecting the edges of the planks on the right side of the shed;

- a second point $\mathbf{v}_2$ is computed by intersecting the edges of the planks on the left side of the shed and the parallel edges on the roof;

- the plane vanishing line $\mathbf{l}$ is computed by joining those two points ($\mathbf{l} = \mathbf{v}_1 \times \mathbf{v}_2$);

- the distance of the top of the frame of the window on the left from the ground has been measured on site and used as reference to compute α as in (9).

- the line $\mathbf{l}_{x'}$, the image of a horizontal line, is selected in the image by choosing any two points on it;

- the associated vanishing point $\mathbf{v}_h$ is computed as $\mathbf{v}_h = \mathbf{l}_{x'} \times \mathbf{l}$;

- the line $\mathbf{l}_x$, which is the image of a line parallel to $\mathbf{l}_{x'}$ in the scene is constrained to pass through $\mathbf{v}_h$, therefore $\mathbf{l}_x$ is specified by choosing one additional point on it;

- a point $\mathbf{x}'$ is selected along the line $\mathbf{l}_{x'}$ and its corresponding point $\mathbf{x}$ on the line $\mathbf{l}_x$ computed as $\mathbf{x} = (\mathbf{x}' \times \mathbf{v}) \times \mathbf{l}_x$;

- equation (10) is now applied to the pair of points $\mathbf{x}$, $\mathbf{x}'$ to compute the distance $Z = 294.3cm$.
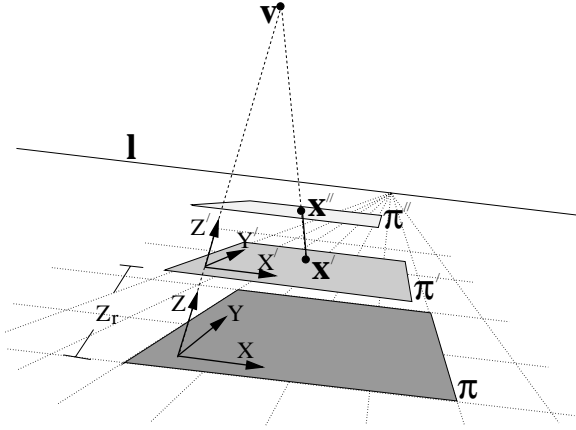
Fig. 11. Measuring the distance between any two planes π' and π'' parallel to the reference plane π.

3.1.2. *Distance between any two parallel planes*

The projection matrix P from the world to the image is defined in (4) with respect to a coordinate frame on the reference plane (fig. 9). In this section we determine the projection matrix P′ referred to the parallel plane π' and we show how distances from the plane π' can be computed.

Suppose the world coordinate system is translated by Z_r from the plane π onto the plane π' along the reference direction (fig. 11), then we can parametrize the new projection matrix P′ as:

$$\mathtt{P}' = \begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_3 & Z_r\mathbf{p}_3 + \mathbf{p}_4 \end{bmatrix} \qquad (13)$$

Note that if $Z_r = 0$ then P′ = P as expected.

The distance Z' of the plane π'' from the plane π' in space can be computed as (*c.f.* (10)).

$$Z' = -\frac{\|\mathbf{x}' \times \mathbf{x}''\|}{\rho'\|\mathbf{p}_3 \times \mathbf{x}''\|} \qquad (14)$$

with

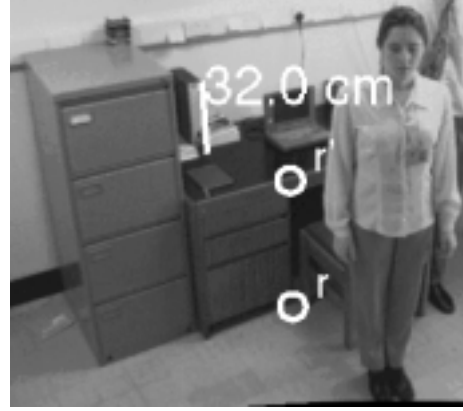$$\rho' = \frac{\mathbf{x}' \cdot \mathbf{p}_4}{1 + Z_r\mathbf{p}_3 \cdot \mathbf{p}_4}$$

Fig. 12. **Measuring heights of objects on separate planes:** The height of the desk is known and the height of the file on the desk is computed.

Worked example. In figure 12 the height of a file on a desk is computed from the height of the desk itself

- The ground is the reference plane π and the top of the desk is the plane denoted as π' in figure 11;

- the plane vanishing line and vertical vanishing point are computed as usual by intersecting parallel edges;

- the distance Z_r between the points $\mathbf{r}$ and $\mathbf{r}'$ is known (the height of the desk has been measured on site) and used to compute the α parameter from (9);

- equation (14) is now applied to the end points of the marked segment to compute the height $Z' = 32.0\,\text{cm}$.

3.2. *Measurements on parallel planes*

As described in section 2.2, given the homology between two planes π and π' in the pencil we can transfer all points from one plane to the other and make affine measurements in either plane.

The homology between the planes can be derived directly from the two projection matrices (4) and (13). The plane-to-image homographies are extracted from the projection matrices ignoring the third column, to give:

$$\mathtt{H} = \begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_4 \end{bmatrix}, \quad \mathtt{H}' = \begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & Z_r\mathbf{p}_3 + \mathbf{p}_4 \end{bmatrix}$$

Then $\tilde{\mathtt{H}} = \mathtt{H}'\mathtt{H}^{-1}$ maps image points on the plane π onto points on the plane π' and so defines the homology.

By inspection, since $\mathbf{p}_1 \cdot \mathbf{p}_4 = 0$ and $\mathbf{p}_2 \cdot \mathbf{p}_4 = 0$ then $(\mathtt{I} + Z_r\mathbf{p}_3\mathbf{p}_4^\top)\mathtt{H} = \mathtt{H}'$, hence the homology
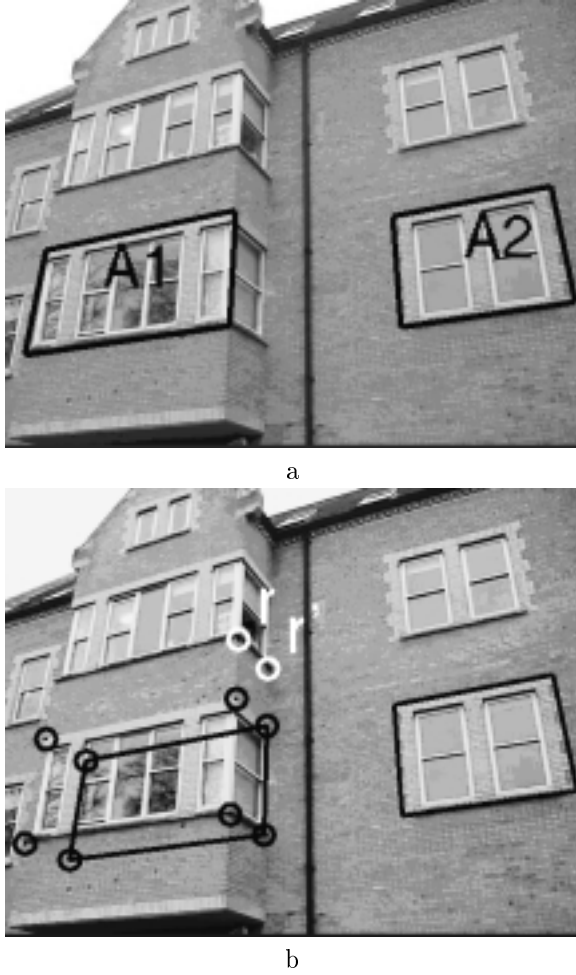
a

b

Fig. 13. **Measuring ratios of areas on separate planes:** (a) original image with two windows hilighted; (b) the left window is transferred onto the plane identified by $\mathbf{r}'$ by the homology mapping (16). The two areas now lie on the same plane and can, therefore, be compared. The ratio between the areas of the two windows is then computed as: $\frac{A_1}{A_2} = 1.45$.

matrix $\tilde{\mathtt{H}}$ is:

$$\tilde{\mathtt{H}} = \mathtt{I} + Z_r \mathbf{p}_3 \mathbf{p}_4^\top \qquad (15)$$

Alternatively from the (4) the homology matrix can be written as:

$$\tilde{\mathtt{H}} = \mathtt{I} + \psi \mathbf{v} \bar{\mathbf{l}}^\top \qquad (16)$$

with $\mathbf{v}$ the vertical vanishing point, $\bar{\mathbf{l}}$ the normalized plane vanishing line and $\psi = \alpha Z_r$ (*c.f.* (3)).

If the distance Z_r and the last two columns of the matrix $\mathtt{P}$ are known then the homology between the two planes $\boldsymbol{\pi}$ and $\boldsymbol{\pi}'$ is computed as in (15). Otherwise, if only $\mathbf{v}$ and $\mathbf{l}$ are known and

two corresponding points $\mathbf{r}$ and $\mathbf{r}'$ are viewed, then the homology parameter ψ in (16) can be computed from (9) (remember that $\alpha Z_r = \psi$) without knowing either the distance Z_r between the two planes or the α parameter.

Examples of homology transfer and affine measurements are shown in figures 8 and 13.

Worked example. In figure 13 we compute the ratio between the areas of two windows $\frac{A_1}{A_2}$ in the world.

- The orthogonal vanishing point $\mathbf{v}$ is computed by intersecting the edges of the small windows linking the two front planes;

- the plane vanishing line $\mathbf{l}$ (common to both front planes) is computed by intersecting two sets of parallel edges on the two planes;

- the only remaining parameter ψ of the homology $\tilde{\mathtt{H}}$ in (16) is computed from (9) as

$$\psi = -\frac{||\mathbf{r} \times \mathbf{r}'||}{(\bar{\mathbf{l}} \cdot \mathbf{r})||\mathbf{v} \times \mathbf{r}'||}$$

- each of the four corners of the window on the left is transferred by the homology $\tilde{\mathtt{H}}$ onto the corresponding points on the plane of the other window (fig. 13b);

Now we have two quadrilaterals on the same plane

- the image is affine-warped pulling the plane vanishing line to infinity [LZ98];

- the ratio between the two areas in the world is computed as the ratio between the areas in the affine-warped image. We obtain $\frac{A_1}{A_2} = 1.45$.

3.3. *Determining camera position*

Suppose the camera centre is $\mathbf{C} = (X_c, Y_c, Z_c, W_c)^\top$ (see figure 2). Then since $\mathtt{P}\mathbf{C} = \mathbf{0}$ we have

$$\mathtt{P}\mathbf{C} = \mathbf{p}_1 X_c + \mathbf{p}_2 Y_c + \mathbf{p}_3 Z_c + \mathbf{p}_4 W_c = \mathbf{0} \qquad (17)$$

The solution to this set of equations is given (using Cramer's rule) by

$$
\begin{aligned}
X_c &= -\det\left[\, \mathbf{p}_2 \;\; \mathbf{p}_3 \;\; \mathbf{p}_4 \,\right], \\
Y_c &= \det\left[\, \mathbf{p}_1 \;\; \mathbf{p}_3 \;\; \mathbf{p}_4 \,\right], \\
Z_c &= -\det\left[\, \mathbf{p}_1 \;\; \mathbf{p}_2 \;\; \mathbf{p}_4 \,\right], \\
W_c &= \det\left[\, \mathbf{p}_1 \;\; \mathbf{p}_2 \;\; \mathbf{p}_3 \,\right] \qquad (18)
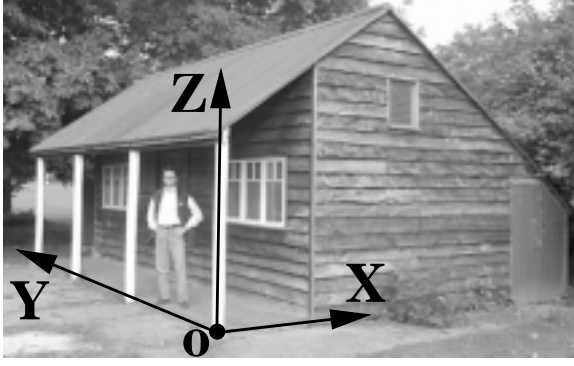\end{aligned}
$$

and the location of the camera centre is defined.

Fig. 14. **Computing the location of the camera:**
Equations (18) are used to obtain:
X_c = -381.0cm, Y_c = -653.7cm, Z_c = 162.8cm.

If α is unknown we can write:

$$
\begin{aligned}
X_c &= -\det \begin{bmatrix} \mathbf{p}_2 & \mathbf{v} & \mathbf{p}_4 \end{bmatrix}, \\
Y_c &= \det \begin{bmatrix} \mathbf{p}_1 & \mathbf{v} & \mathbf{p}_4 \end{bmatrix}, \\
\alpha Z_c &= -\det \begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_4 \end{bmatrix}, \\
W_c &= \det \begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{v} \end{bmatrix}
\end{aligned}
\tag{19}
$$

and we obtain the distance Z_c of the camera centre from the plane up to the affine scale factor α. As before, we may upgrade the distance Z_c to metric with knowledge of α, or use knowledge of the camera height to compute α and upgrade the affine structure.

Note that affine viewing conditions (where the camera centre is at infinity) present no problem in expressions (18) and (19), since in this case we have $\bar{\mathbf{l}} = \begin{bmatrix} 0 & 0 & * \end{bmatrix}^{\top}$ and $\mathbf{v} = \begin{bmatrix} * & * & 0 \end{bmatrix}^{\top}$. Hence $W_c = 0$ so we obtain a camera centre on the plane at infinity, as expected. This point on $\boldsymbol{\pi}_{\infty}$ represents the viewing direction for the parallel projection.

If the viewpoint is finite (i.e. not affine viewing conditions) then the formula for αZ_c may be developed further by taking the scalar product of both sides of (17) with the vanishing line $\bar{\mathbf{l}}$. The result is

$$
\alpha Z_c = -\frac{1}{\bar{\mathbf{l}} \cdot \mathbf{v}}
\tag{20}
$$

Worked example. In figure 14 the position of the camera centre with respect to the chosen Cartesian coordinates system is determined.

Note that in this case we have chosen $\mathbf{p}_4$ to be the point $\mathbf{o}$ in the figure instead of $\bar{\mathbf{l}}$.

- The ground plane (X, Y plane) is the reference;

- the vertical vanishing point is computed by intersecting vertical edges;

- the two sides of the rectangular base of the porch have been measured thus providing the position of four points on the reference plane. The world-to-image homography is computed from those points [CRZ99a];

- the distance of the top of the frame of the window on the left from the ground has been measured on site and used as reference to compute α as in (9).

- the 3D position of the camera centre is then computed simply by applying equations (18). We obtain

$$X_c = -381.0cm \;\; Y_c = -653.7cm \;\; Z_c = 162.8cm$$

In figure 22c, the camera has been superimposed into a virtual view of the reconstructed scene.

4. Uncertainty Analysis

Feature detection and extraction – whether manual or automatic (e.g. using an edge detector) – can only be achieved to a finite accuracy. Any features extracted from an image, therefore, are subject to measurements errors. In this section we consider how these errors propagate through the measurement formulae in order to quantify the uncertainty on the final measurements [Fau93]. This is achieved by using a first order error analysis.

We first analyse the uncertainty on the projection matrix and then the uncertainty on distance measurements.

4.1. *Uncertainty on the P matrix*

The uncertainty in P depends on the location of the vanishing line, the location of the vanishing point, and on α, the affine scale factor. Since only the final two columns contribute, we model the uncertainty in P as a 6×6 homogeneous covariance matrix, Λ_{P}. Since the two columns have only five degrees of freedom (two for $\mathbf{v}$, two for $\mathbf{l}$ and one for α), the covariance matrix is singular, with rank five.

Assuming statistical independence between the two column vectors $\mathbf{p}_3$ and $\mathbf{p}_4$ the 6×6 rank five covariance matrix $\Lambda_\mathbf{P}$ can be written as:

$$\Lambda_\mathbf{P} = \begin{pmatrix} \Lambda_{\mathbf{p}_3} & 0 \\ 0 & \Lambda_{\mathbf{p}_4} \end{pmatrix} \qquad (21)$$

Furthermore, assuming statistical independence between α and $\mathbf{v}$, since $\mathbf{p}_3 = \alpha \mathbf{v}$, we have:

$$\Lambda_{\mathbf{p}_3} = \alpha^2 \Lambda_\mathbf{v} + \sigma_\alpha^2 \mathbf{v} \mathbf{v}^\top \qquad (22)$$

with $\Lambda_\mathbf{v}$ the homogeneous 3×3 covariance of the vanishing point $\mathbf{v}$ and the variance σ_α^2 computed as in Appendix D.

Since $\mathbf{p}_4 = \bar{\mathbf{l}} = \frac{\mathbf{l}}{||\mathbf{l}||}$ its covariance is:

$$\Lambda_{\mathbf{p}_4} = \frac{\partial \mathbf{p}_4}{\partial \mathbf{l}} \Lambda_\mathbf{l} {\frac{\partial \mathbf{p}_4}{\partial \mathbf{l}}}^\top \qquad (23)$$

where the 3×3 Jacobian $\frac{\partial \mathbf{p}_4}{\partial \mathbf{l}}$ is

$$\frac{\partial \mathbf{p}_4}{\partial \mathbf{l}} = \frac{\mathbf{l} \cdot \mathbf{l} \mathbf{I} - \mathbf{l} \mathbf{l}^\top}{(\mathbf{l} \cdot \mathbf{l})^{\frac{3}{2}}}$$

4.2. Uncertainty on measurements between planes

When making measurements between planes (10), uncertainty arises from the uncertain image locations of the points $\mathbf{x}$ and $\mathbf{x}'$ and from the uncertainty in $\mathbf{P}$.

The uncertainty in the end points $\mathbf{x}$, $\mathbf{x}'$ of the length to be measured (resulting largely from the finite accuracy with which these features may be located in the image) is modeled by covariance matrices $\Lambda_\mathbf{x}$ and $\Lambda_{\mathbf{x}'}$.

4.2.1. Maximum likelihood estimation of the end points and uncertainties.
In this section we assume a noise-free $\mathbf{P}$ matrix. This assumption will be removed in section 4.2.2.

Since in the error-free case, $\mathbf{x}$ and $\mathbf{x}'$ must be aligned with the vertical vanishing point we can determine the maximum likelihood estimates ($\hat{\mathbf{x}}$ and $\hat{\mathbf{x}}'$) of their true locations by minimizing the sum of the Mahalanobis distances between the input points $\mathbf{x}$ and $\mathbf{x}'$ and their MLE estimates $\hat{\mathbf{x}}$ and $\hat{\mathbf{x}}'$

$$\min_{\hat{\mathbf{x}}_2, \hat{\mathbf{x}}'_2,} \quad [\quad (\mathbf{x}_2 - \hat{\mathbf{x}}_2)^\top \Lambda_{\mathbf{x}_2}^{-1} (\mathbf{x}_2 - \hat{\mathbf{x}}_2) \qquad (24)$$
$$+ \; (\mathbf{x}'_2 - \hat{\mathbf{x}}'_2)^\top \Lambda_{\mathbf{x}'_2}^{-1} (\mathbf{x}'_2 - \hat{\mathbf{x}}'_2) \quad]$$
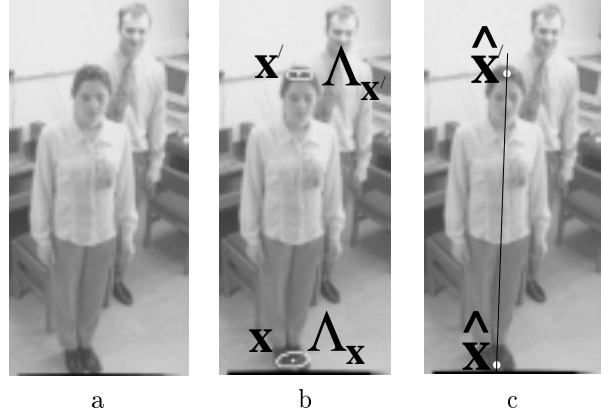

Fig. 15. **Maximum likelihood estimation of the end points:** (a) Original image (closeup of fig. 16b). (b) The uncertainty ellipses of the end points, $\Lambda_\mathbf{x}$ and $\Lambda_{\mathbf{x}'}$, are shown. These ellipses are defined manually, and indicate a confidence region for localizing the points. (c) MLE end points $\hat{\mathbf{x}}$ and $\hat{\mathbf{x}}'$ are aligned with the vertical vanishing point (outside the image).

subject to the *alignment constraint*

$$\mathbf{v} \cdot (\hat{\mathbf{x}} \times \hat{\mathbf{x}}') = 0 \qquad (25)$$

(the subscript 2 indicates inhomogeneous 2-vectors).

This is a constrained minimization problem. A closed-form solution can be found (by the Lagrange multiplier method) in the special case that

$$\Lambda_{\mathbf{x}'_2} \; = \; \gamma^2 \; \Lambda_{\mathbf{x}_2}$$

with γ a scalar, but, unfortunately, in the general case there is no closed-form solution to the problem. Nevertheless, in the general case, an initial solution can be computed by using the approximation given in Appendix B and then refining it by running a numerical algorithm such as Levenberg-Marquardt.

Once the MLE end points have been estimated, we use standard techniques [Fau93, Cla98] to obtain a first order approximation to the 4×4, rank-three covariance of the MLE 4-vector $\hat{\zeta}^\top = (\; \hat{\mathbf{x}}'^\top_2 \quad \hat{\mathbf{x}}^\top_2 \;)$. Figure 15 illustrates the idea (see Appendix C for details).

4.2.2. Uncertainty on distance measurements.
Assuming noise in both end points and in the projection matrix, and statistical independence between $\hat{\zeta}$ and $\mathbf{P}$ we obtain a first order approximation for the variance of the distance Z of a point

from a plane:

$$\sigma_Z^2 = \boldsymbol{\nabla}_Z \begin{pmatrix} \Lambda_{\hat{\zeta}} & 0 \\ 0 & \Lambda_{\mathtt{P}} \end{pmatrix} \boldsymbol{\nabla}_Z^{\top} \qquad (26)$$

where $\boldsymbol{\nabla}_Z$ is the 1×10 Jacobian matrix of the function (10) which maps the projection matrix and the end points $\mathbf{x}$, $\mathbf{x}'$ to their world distance Z. The computation of $\boldsymbol{\nabla}_Z$ is explained in detail in Appendix C.

4.3.   Uncertainty on camera position

The distance of the camera centre from the reference plane is computed according to (20) which can be rewritten as:

$$Z_c = -(\mathbf{p}_4 \cdot \mathbf{p}_3)^{-1} \qquad (27)$$

If we assume an exact $\mathtt{P}$ matrix, then the camera distance is exact too, in fact it depends only on the matrix elements of $\mathtt{P}$. Likewise, the accuracy of Z_c depends only on the accuracy of the $\mathtt{P}$ matrix.

Equation (27) maps $\mathcal{R}^6$ into $\mathcal{R}$, and the associated 1×6 Jacobian matrix $\boldsymbol{\nabla} Z_c$ is readily derived to be

$$\boldsymbol{\nabla} Z_c = Z_c^2 \left(\mathbf{p}_4^{\top} \;\; \mathbf{p}_3^{\top} \right)$$

and, from a first order analysis the variance of Z_c is

$$\sigma_{Z_c}^2 = \boldsymbol{\nabla} Z_c \Lambda_{\mathtt{P}} \boldsymbol{\nabla} Z_c^{\top} \qquad (28)$$

where $\Lambda_{\mathtt{P}}$ is computed in section 4.1.

The variances $\sigma_{X_c}^2$ and $\sigma_{Y_c}^2$ of the X, Y location of the camera can be comamuted in a similar way [CRZ99a].

4.4.   Example - uncertainty on measurements between planes

In this section we show the effects of the number of reference distances and image localization error on the predicted uncertainty in measurements.

An image obtained from a security camera with a poor quality lens is shown in figure 16a. It has been corrected for radial distortion using the method described by Devernay and Faugeras [DF95], and the floor taken as the reference plane.

The scene is calibrated by identifying two points $\mathbf{v}_1, \mathbf{v}_2$ on the reference plane's vanishing line

(shown in white at the top of each image) and the vertical vanishing point $\mathbf{v}$. These points are computed by intersecting sets of parallel lines. The uncertainty on each point is assumed to be Gaussian and isotropic with standard deviation $0.1 pixels$. The uncertainty of the vanishing line is derived from a first order propagation through the vector product operation $\mathbf{l} = \mathbf{v}_1 \times \mathbf{v}_2$. The projection matrix $\mathtt{P}$ is therefore uncertain with its covariance given by (21).

In addition the end points of the height to be measured are assumed to be uncertain and their covariances estimated as in section 4. The uncertainties in the height measurements shown are computed as 3-standard deviation intervals.

In figure 16b one reference height is used to compute the affine scale factor α from (9) (i.e. the minimum number of references). Uncertainty has been assumed in the reference heights, vertical vanishing point and plane vanishing line. Once α is computed other measurements in the same direction are metric. The height of the man has been computed and shown in the figure. It differs by $4mm$ from the known true value.

The uncertainty associated with the height of the man is computed from (26) and displayed in figure 16b. Note that the true height value falls always within the computed 3-standard deviation range as expected.

As the number of reference distances is increased (see figures 16c,d), so the uncertainty on $\mathtt{P}$ (in fact just on α) decreases, resulting in a decrease in uncertainty of the measured height, as theoretically expected (see Appendix D). Equation (12) has been employed, here, to metric-calibrate the distance from the floor.

Figure 17 shows images of the same scene with the same people, but acquired from a different point of view. As before the uncertainty on the measurements decreases as the number of references increases (fig. 17b,c). The measurement is the same as in the previous view (fig. 16) thus demonstrating invariance to camera location.

Figure 18 shows an example, where the height of the woman and the related uncertainty are computed for two different orientations of the uncertainty ellipses of the end points. In figure 18b the two input ellipses of figure 18a have been rotated by an angle of approximately 40°, maintaining the size and position of the centres. The angle be-
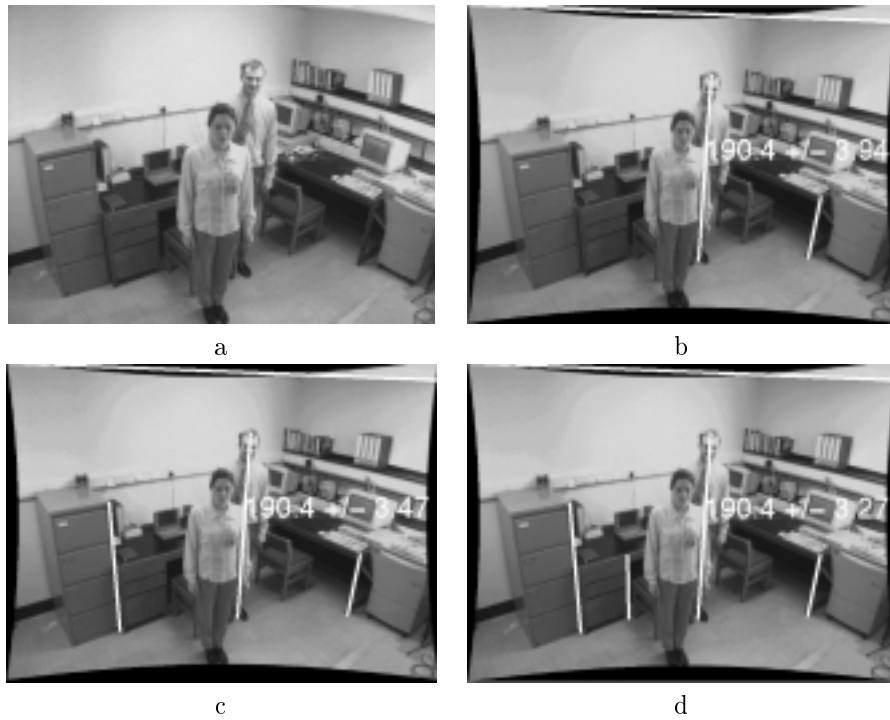
Fig. 16.   **Measuring heights and estimating their uncertainty:** (a) Original image; (b) Image corrected for radial distortion and measurements superimposed. With only *one* supplied reference height the man's height has been measured to be Z = 190.4 ± 3.94cm, (*c.f.* ground truth value 190cm). The uncertainty has been estimated by using (26) (the uncertainty bound is at ± 3 std.dev.). (c) With *two* reference heights Z = 190.4 ± 3.47cm. (d) With *three* reference heights Z = 190.4 ± 3.27cm. Note that in the limit $\Lambda_P = 0$ (error-free P matrix) the height uncertainty reduces to 2.16cm for all (b,c,d); the residual error, in this case, is due only to the error on the two end points.
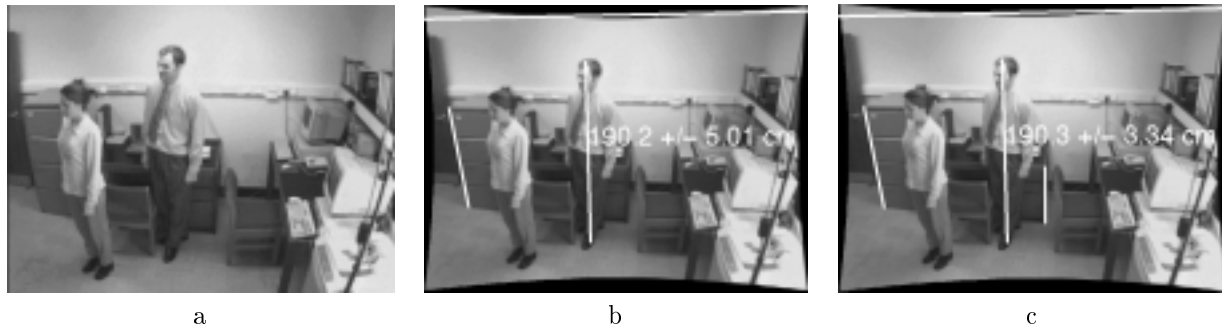

Fig. 17.   **Measuring heights and estimating their uncertainty, second point of view:** (a) Original image; (b) the image has been corrected for radial distortion and height measurements computed and superimposed. With *one* supplied reference height Z = 190.2 ± 5.01cm (*c.f.* ground truth value 190cm). (c) With *two* reference heights Z = 190.4 ± 3.34cm. See figure 16 for details.

tween the direction defined by the major axes (direction of maximum uncertainty) of each ellipse and the measuring direction is smaller than in fig. 18a and the uncertainty in the measurements greater as expected.

4.5.   Monte Carlo test

In this section we validate the first order error analysis described above by computing the uncertainty of the height of the man in fig. 16d using our first order analytical method and comparing
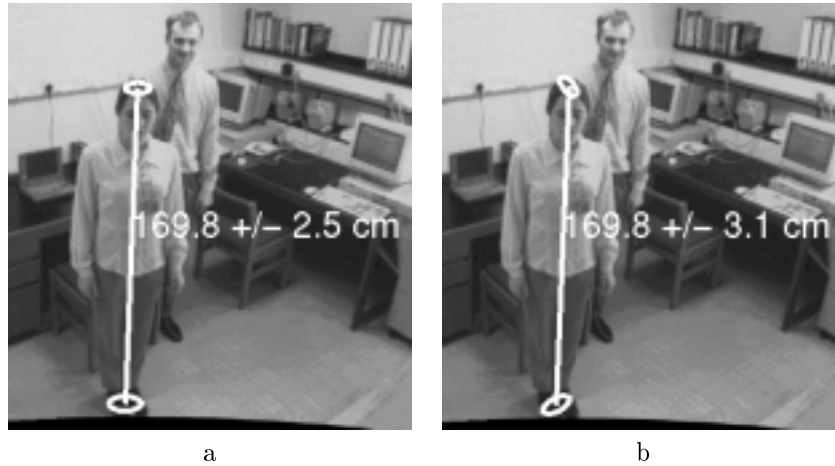
a   b

Fig. 18. **Estimating the uncertainty in height measurements for different orientations of the input 3-standard deviation uncertainty ellipses:** (a) Cropped version of image 16a with measurements superimposed: Z = 169.8 ± 2.5cm (at 3-standard deviations). The ground truth is Z = 170cm, it lies within the computed range. (b) the input ellipses have been rotated keeping their size and position fixed: Z = 169.8 ± 3.1cm (at 3-standard deviations). The height measurement is less accurate.
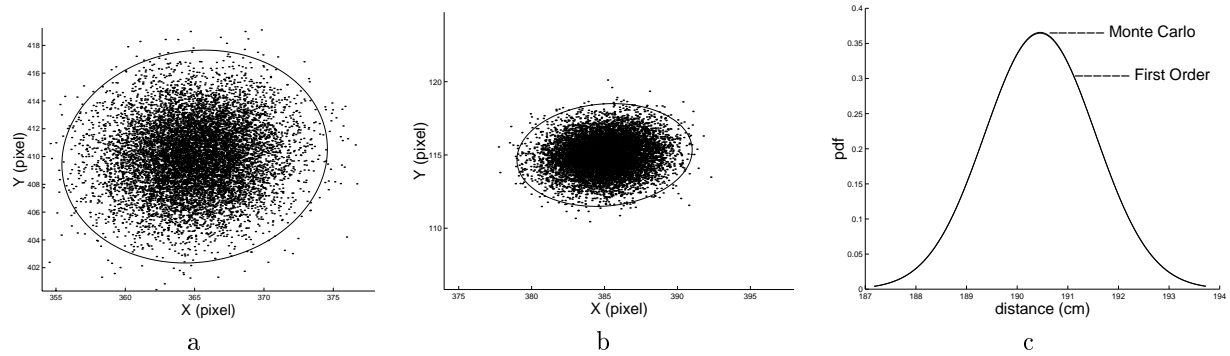

a   b   c

Fig. 19. **Monte Carlo simulation of the example in fig. 16d:** (a) distribution of the input base point **x** and the corresponding 3-standard deviation ellipse. (b) distribution of the input top point **x**′ and the corresponding 3-standard deviation ellipse. Note that figures (a) and (b) are drawn at the same scale. (c) the analytical and simulated distributions of the computed distance Z. The two curves are almost perfectly overlapping.

it to the uncertainty derived from Monte Carlo simulations as described in table 1.

Specifically, we compute the statistical standard deviation of the man's height from a reference plane and compare it with the standard deviation obtained from the first order error analysis.

Uncertainty is modeled as Gaussian noise and described by covariance matrices. We assume noise on the end points of the three reference distances. Uncertainty is assumed also on the vertical vanishing point, the plane vanishing line and on the end points of the height to be measured.

Figure 19 shows the results of the test. The base point is randomly distributed according to a 2D non-isotropic Gaussian about the mean location **x** (on the feet of the man in figure 16) with covariance matrix $\Lambda_{\mathbf{x}}$ (figure 19a). Similarly the top point is randomly distributed according to a 2D non-isotropic Gaussian about the mean location **x**′ (on the head of the man in figure 16), with covariance $\Lambda_{\mathbf{x}'}$ (figure 19b).

The two covariance matrices are respectively:

$$\Lambda_{\mathbf{x}} = \begin{pmatrix} 10.18 & 0.59 \\ 0.59 & 6.52 \end{pmatrix} \quad \Lambda_{\mathbf{x}'} = \begin{pmatrix} 4.01 & 0.22 \\ 0.22 & 1.36 \end{pmatrix}$$

Table 1. Monte Carlo simulation.

- for j=1 to S (with S = number of samples)
  - For each reference: given the measured reference end points $\mathbf{r}$ (on the reference plane) and $\mathbf{r}'$, generate a random base point $\mathbf{r}_j$, a random top point $\mathbf{r}'_j$ and a random reference distance $\mathbf{Z}_{r_j}$ according to the associated covariances.
  - Generate a random vanishing point according to its covariance $\Lambda_{\mathbf{v}}$.
  - Generate a random plane vanishing line according to its covariance $\Lambda_{\mathbf{l}}$.
  - Compute the α parameter by applying (12) to the references, and the current P matrix (4).
  - Generate a random base point $\mathbf{x}_j$ and a random top point $\mathbf{x}'_j$ for the distance to be computed according to their respective covariances $\Lambda_{\mathbf{x}}$ and $\Lambda_{\mathbf{x}'}$.
  - Project the points $\mathbf{x}_j$ and $\mathbf{x}'_j$ onto the best fitting line through the vanishing point (see sect. 4.2.1).
  - Compute the current distance Z_j by applying (10).
- The statistical standard deviation of the population of simulated Z_j values is computed as

$$\sigma'^2_Z = \frac{\sum_{j=1}^S (Z_j - \overline{Z})^2}{S}$$

and compared to the analytical one (26).

Suitable values for the covariances of the three references, the vanishing point and the vanishing line have been used. The simulation has been run with $S = 10000$ samples.

Analytical and simulated distributions of Z are plotted in figure 19c; the two curves are almost overlapping. Slight differences are due to the assumptions of statistical independence (21, 22, 26) and first order truncation introduced by the error analysis.

A comparison between statistical and analytical standard deviations is reported in the table below with the corresponding relative error:

First Order σ_Z	Monte Carlo σ'_Z	relative error $\frac{\|\sigma_Z - \sigma'_Z\|}{\sigma'_Z}$
1.091 cm	1.087 cm	0.37 %

Note that $Z = 190.45 cm$ and the associated first order uncertainty $3 * \sigma_Z = 3.27 cm$ is shown in fig. 16d.

In the limit $\Lambda_{\mathsf{P}} = 0$ (error-free P matrix) the simulated and analytical results are even closer.

This result shows the validity of the first order approximation in this case and numerous other examples have followed the same pattern. However some care must be exercised since as the input uncertainty increases, not only does the output uncertainty increases, but the relative error between statistical and analytical output standard deviations also increases. For large covariances, the assumption of linearity and therefore the first order analysis no longer holds.

This is illustrated in the table below where the relative error is shown for various increasing values of the input uncertainties. The uncertainties of references distances and end points are multiplied by the increasing factor γ; for instance, if $\Lambda_{\mathbf{x}}$ is the covariance of the image point $\mathbf{x}$ then $\Lambda_{\mathbf{x}}(\gamma) = \gamma^2 \Lambda_{\mathbf{x}}$.

γ	1	5	10	20	30
$\frac{\|\sigma_Z - \sigma'_Z\|}{\sigma'_Z}$ (%)	0.37	1.68	3.15	8.71	16.95

In the *affine case* (when the vertical vanishing point and the plane vanishing line are at infinity) the first order error propagation is exact (no longer just an approximation as in the general projective case), and the analytic and Monte Carlo simulation results coincide.

5. Applications

5.1. Forensic science

A common requirement in surveillance images is to obtain measurements from the scene, such as the height of a felon. Although, the felon has usually departed the scene, reference lengths can be measured from fixtures such as tables and windows.

In figure 20 we compute the height of the suspicious person standing next to the phonebox. The ground is the reference plane and the vertical is the reference direction. The edges of the paving stones are used to compute the plane vanishing line, the edges of the phonebox to compute the vertical vanishing point; and the height of the phonebox provides the metric calibration in the vertical direction (figure 20b). The height of the
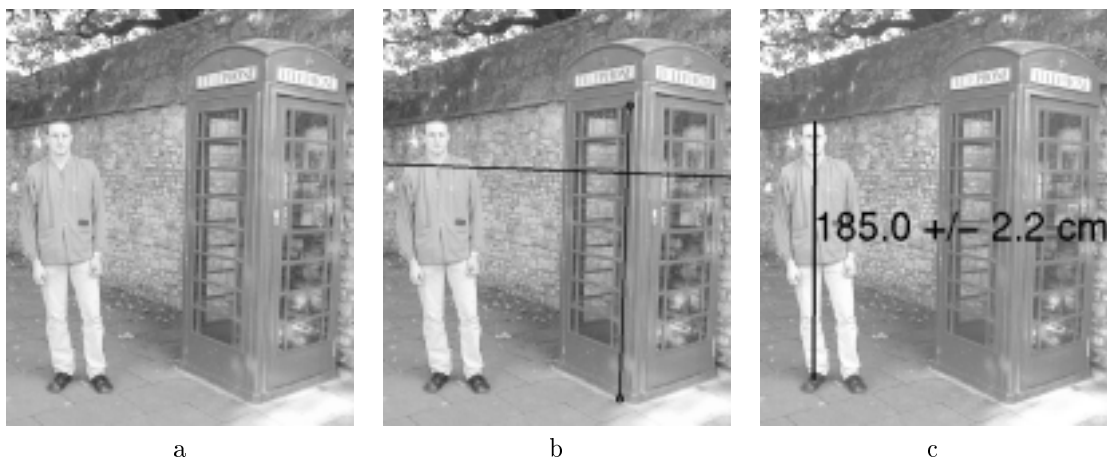
Fig. 20. **The height of a person standing by a phonebox is computed:** (a) Original image. (b) The ground plane is the reference plane, and its vanishing line is computed from the paving stones on the floor. The vertical vanishing point is computed from the edges of the phonebox, whose height is known and used as reference. Vanishing line and reference height are shown. (c) The computed height of the person and the estimated uncertainty are shown. The veridical height is 187cm. Note that the person is leaning slightly on his right foot.
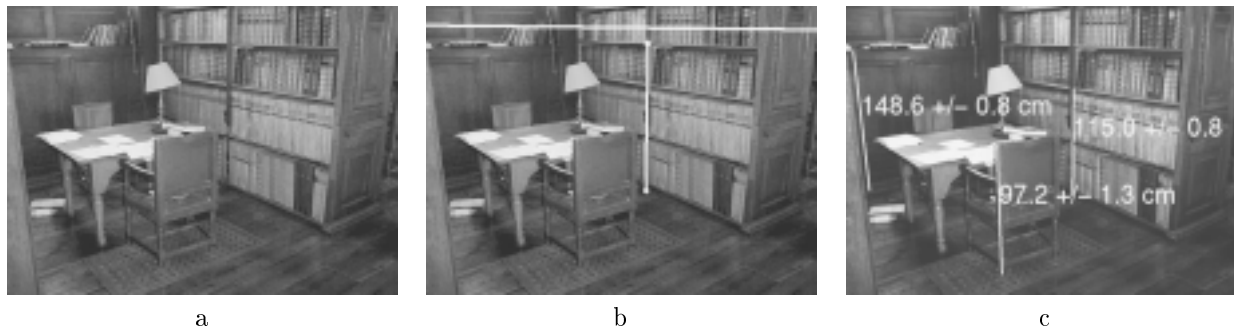

Fig. 21. **Measuring height of furniture in The Queen's College Upper Library, Oxford:** (a) Original image. (b) The plane vanishing line (white horizontal line) and reference height (white vertical line) are superimposed on the original image; the marked shelf is 156cm high. (c) Computed heights and related uncertainties; the uncertainty bound is at $\pm$ 3 std.dev. The ground truth is: 115cm for the right hand shelf, 97cm for the chair and 149cm for the shelf at the left. Note that the ground truth always falls within the computed uncertainty range.

person is then computed using (10) and shown in figure 20c. The ground truth is 187cm, note that the person is leaning slightly down on his right foot.

The associated uncertainty has also been estimated; two uncertainty ellipses have been defined, one on the head of the person and one on the feet and then propagated through the chain of computations as described in section 4 to give the 2.2cm 3-standard deviation uncertainty range shown in figure 20c.

5.2. Furniture measurements

In this section another application is described. Heights of furniture like shelves, tables or windows in an indoor environment are measured.

Figure 21a shows a desk in The Queen's College upper library in Oxford. The floor is the reference plane and its vanishing line has been computed by intersecting edges of the floorboards. The vertical vanishing point has been computed by intersecting the vertical edges of the bookshelf. The vanishing line is shown in figure 21b with the reference height used. Only one reference height (minimal set) has been used in this example.
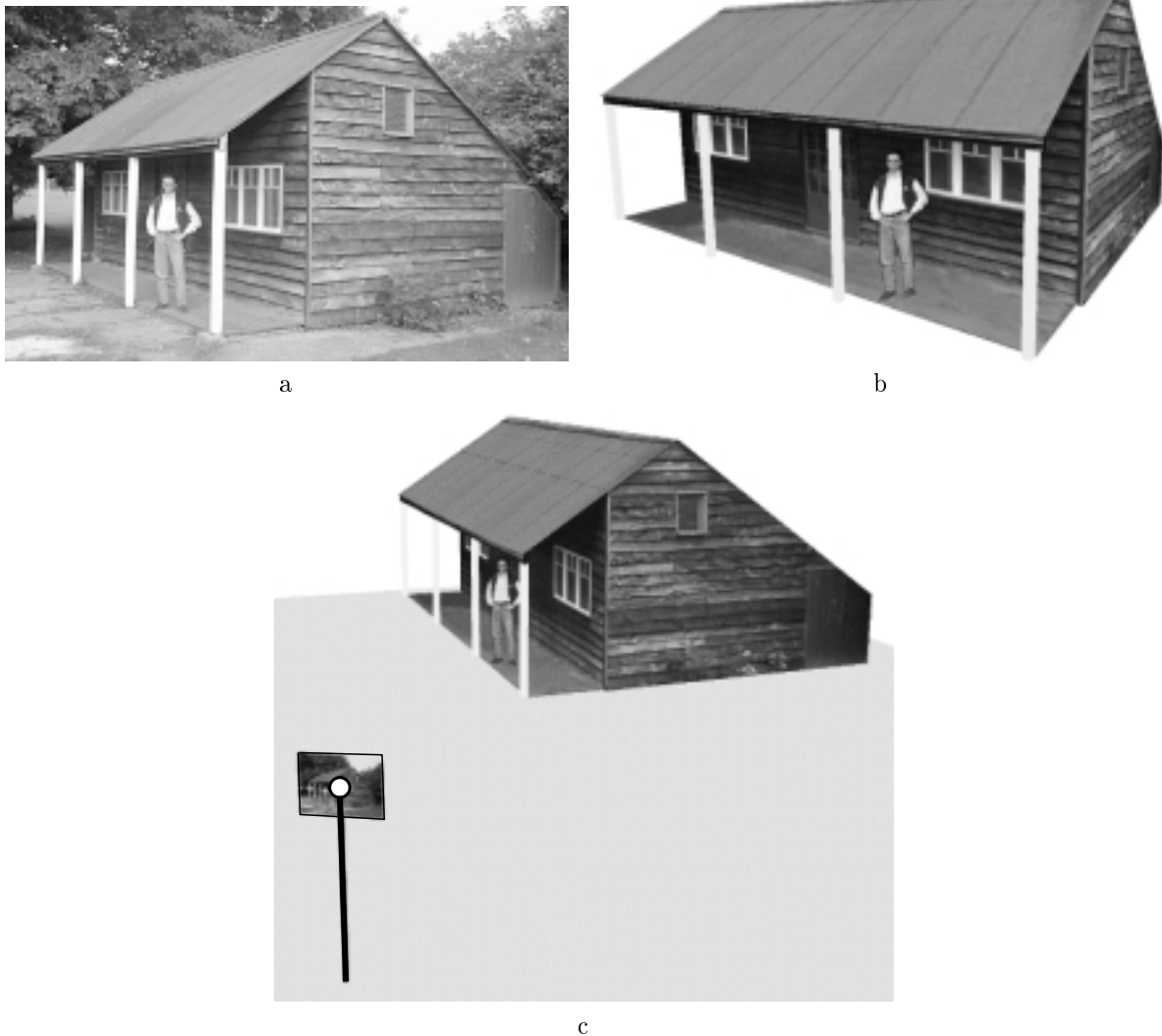
Fig. 22. **Complete 3D reconstruction of a real scene:** (a) original image; (b) a view of the reconstructed 3D model; (c) A view of the reconstructed 3D model which shows the position of the camera centre (plane location X,Y and height) with respect to the scene.

The computed heights and associated uncertainties are shown in figure 21c. The uncertainty bound is ±3 standard deviations. Note that the ground truth always falls within the computed uncertainty range. The height of the camera is computed as $1.71m$ from the floor.

5.3.   *Virtual modelling*

In figure 22 we show an example of complete 3D reconstruction of a real scene from a single image. Two sets of horizontal edges are used to compute the vanishing line for the ground plane, and vertical edges used to compute the vertical vanishing point.

The distance of the top of the window to the ground, and the height of one of the pillars are

used as reference heights. Furthermore the two sides of the base of the porch have been measured thus defining the metric calibration of the ground plane.

Figure 22b shows a view of the reconstructed model. Notice that the person is represented simply as a flat silhouette since we have made no attempt to recover his volume. The position of the camera centre is also estimated and superimposed on a different view of the 3D model in figure 22c.

5.4. Modelling paintings

Figure 23 shows a masterpiece of Italian Renaissance painting, "La Flagellazione di Cristo" by Piero della Francesca (1416 - 1492). The painting faithfully follows the geometric rules of perspective, and therefore the methods developed here can be applied to obtain a 3D reconstruction of the scene.

Unlike other techniques [HAA97] whose main aim is to create convincing new views of the painting regardless of the correctness of the 3D geometry, here we reconstruct a geometrically correct 3D model of the viewed scene (see fig. 23c,d).

In the painting analysed here, the ground plane is chosen as reference and its vanishing line computed from the several parallel lines on it. The vertical vanishing point follows from the vertical lines and consequently the relative heights of people and columns can be computed. Figure 23b shows the painting with height measurements superimposed. Christ's height is taken as reference and the heights of the other people are expressed as relative percentage differences. Note the consistency between the height of the people in the foreground with the height of the people in the background.

By assuming a square floor pattern the ground plane has been rectified and the position of each object estimated [LCZ99, CRZ99a, SM99]. The scale of floor relative to heights is set from the ratio between height and base of the frontoparallel archway. The measurements, up to an overall scale factor are used to compute a three dimensional VRML model of the scene.

Figure 23c shows a view of the reconstructed model. Note that the people are represented as flat silhouettes and the columns have been ap-

proximated with cylinders. The partially seen ceiling has been reconstructed correctly. Figure 23d shows a different view of the reconstructed model, where the roof has been removed to show the relative position of the people in the scene.

6.   Summary and Conclusions

We have explored how the affine structure of three-dimensional space may be partially recovered from perspective images in terms of a set of planes parallel to a reference plane and a reference direction not parallel to the reference plane.

Algorithms have been described to obtain different kinds of measurements: measuring the distance between planes parallel to a reference plane; computing area and length ratios on two parallel planes; computing the camera's location.

A first order error propagation analysis has been performed to estimate uncertainties on the projection matrix and on measurements of point or camera location in the space. The error analysis has been validated by using Monte Carlo statistical tests.

Examples have been provided to show the computed measurements and uncertainties on real images.

More generally, affine three-dimensional space may be represented entirely by sets of parallel planes and directions [Ber87]. We are currently investigating how this full geometry is best represented and computed from a single perspective image.

6.1.   Missing base point

A restriction of the measurement method we have presented is the need to identify corresponding points between planes. One case where the method does not apply therefore is that of measuring the distance of a general 3D point to a reference plane (the corresponding point on the reference plane is undefined). Here the homology is under-determined.

One case of interest is when only one view is provided and a light-source casts shadows onto the reference plane. The light-source provides restrictions analogous to a second viewpoint [RF93, RZ96, RN98, VGPZ98], so the projection (in the

a

b

c

d
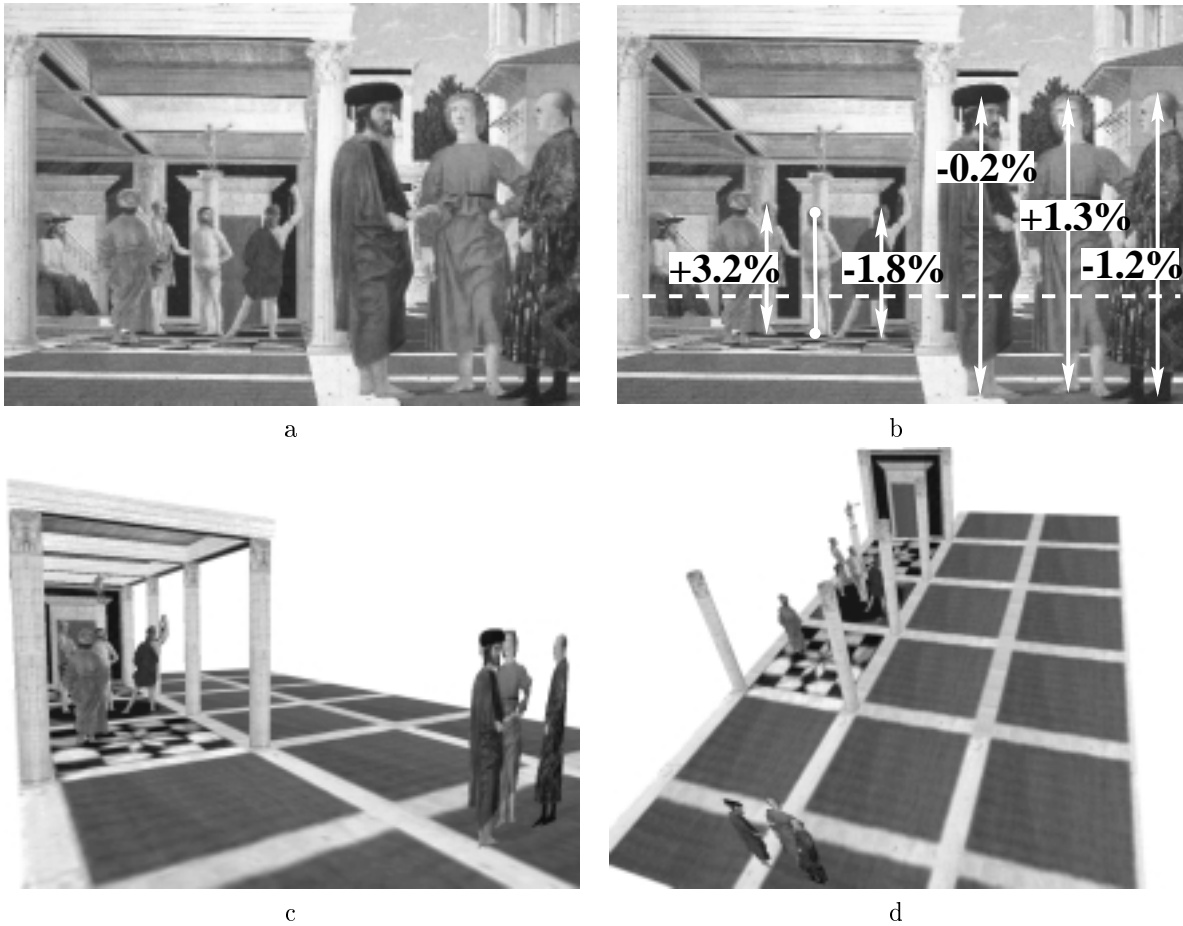
Fig. 23. **Complete 3D reconstruction of a Renaissance painting:** (a) *La Flagellazione di Cristo*, (1460, Urbino, Galleria Nazionale delle Marche). (b) Height measurements are superimposed on the original image. Christ's height is taken as reference and the heights of all the other people are expressed as percent differences. The vanishing line is dashed. (c) A view of the reconstructed 3D model. The patterned floor has been reconstructed in areas where it is occluded by taking advantage of the symmetry of its pattern. (d) Another view of the model with the roof removed to show the relative positions of people and architectural elements in the scene. Note the repeated geometric pattern on the floor in the area delimited by the columns (barely visible in the painting). Note that the people are represented simply as flat silhouettes since it is not possible to recover their volume from one image, they have been cut out manually from the original image. The columns have been approximated with cylinders.

reference direction) of the 3D point onto the reference plane may be determined by making use of the homology defined by the 3D points and their shadows.

Appendix A
Implementation details

Edge detection

Straight line segments are detected by Canny edge detection at subpixel accuracy [Can86]; edge linking; segmentation of the edgel chain at high curvature points; and finally straight line fitting by orthogonal regression to the resulting chain seg-
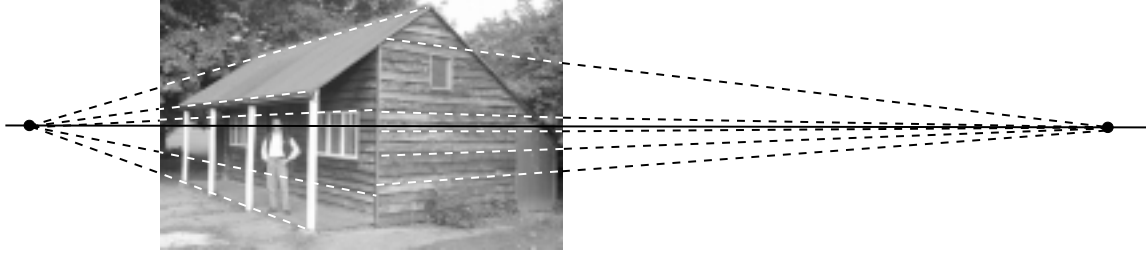
Fig. 24. **Computing the plane vanishing line:** The vanishing line for the reference plane (ground) is shown in solid black. The planks on both sides of the shed define two sets of lines parallel to the ground (dashed); they intersect in points on the vanishing line.

ments (fig. 25b). Lines which are projection of a physical edge in the world often appear broken in the image because of occlusions. A simple merging algorithm based on orthogonal regression has been implemented to merge manually selected edges together. Merging aligned edges to create longer ones increases the accuracy of their location and orientation. An example is shown in figure 25c.

Scene calibration

Vanishing line and vanishing points can be estimated directly from the image and *no* explicit knowledge of the relative geometry between camera and viewed scene is required. Vanishing lines and vanishing points may lie outside the physical image (see fig. 5), but this does not affect the computations.

Computing the vanishing point. All world lines parallel to the *reference direction* are imaged as lines which intersect in the same vanishing point (see fig 2) [Bar83, CT90]. Therefore two such lines are sufficient to define it. However, if more than two lines are available a Maximum Likelihood Estimate algorithm [LZ98] is employed to estimate the point.

Computing the vanishing line. Images of lines parallel to each other and to a plane intersect in points on the plane vanishing line. Therefore two sets of those lines with different directions are sufficient to define the plane vanishing line (fig. 24).

If more than two orientations are available then the computation of the vanishing line is performed by employing a Maximum Likelihood algorithm.

Appendix B
Maximum likelihood estimation of end points for isotropic uncertainties

Given two points $\mathbf{x}$ and $\mathbf{x}'$ with distributions $\Lambda_{\mathbf{x}}$ and $\Lambda_{\mathbf{x}'}$ isotropic but not necessarily equal, we estimate the points $\hat{\mathbf{x}}$ and $\hat{\mathbf{x}}'$ such that the cost function (24) is minimized and the alignment constraint (25) satisfied. It is a constrained minimization problem; a closed form solution esists in this case.

The 2×2 covariance matrices $\Lambda_{\mathbf{x}}$ and $\Lambda_{\mathbf{x}'}$ for the two inhomogeneous end points $\mathbf{x}$ and $\mathbf{x}'$ define two circles with radius $r = \sigma_x = \sigma_y$ and $r' = \sigma_{x'} = \sigma_{y'}$ respectively.

The line $\mathbf{l}$ through the vanishing point $\mathbf{v}$ that best fits the points $\mathbf{x}$ and $\mathbf{x}'$ can be computed as:

$$
\mathbf{l} = \begin{pmatrix} 1 + \sqrt{1 + \xi^2} \\ \xi \\ -(1 + \sqrt{1 + \xi^2})v_x - \xi v_y \end{pmatrix}
$$

with

$$
\xi = 2 \frac{r' d_x d_y + r d'_x d'_y}{r'(d_x^2 - d_y^2) + r(d_x'^2 - d_y'^2)}
$$

where $\mathbf{d}$ and $\mathbf{d}'$ are the following 2-vectors:

$$
\mathbf{d} = \mathbf{x} - \mathbf{v} \quad \mathbf{d}' = \mathbf{x}' - \mathbf{v}
$$

Note that this formulation is valid if $\mathbf{v}$ is finite.

The orthogonal projections of the points $\mathbf{x}$ and $\mathbf{x}'$ onto the line $\mathbf{l}$ are the two estimated homoge-

a

b

c

Fig. 25. **Computing and merging straight edges:** (a) original image; (b) computed edges: some of the edges detected by the Canny edge detector; straight lines have been fitted to them. (c) edges after merging: different pieces of broken lines, belonging to the same edge in space, have been merged together.

neous points $\hat{\mathbf{x}}$ and $\hat{\mathbf{x}}'$:

$$\hat{\mathbf{x}} = \begin{pmatrix} l_y(\mathbf{x} \cdot \mathbf{Fl}) - l_x l_w \\ -l_x(\mathbf{x} \cdot \mathbf{Fl}) - l_y l_w \\ l_x^2 + l_y^2 \end{pmatrix} \quad (29)$$

$$\hat{\mathbf{x}}' = \begin{pmatrix} l_y(\mathbf{x}' \cdot \mathbf{Fl}) - l_x l_w \\ -l_x(\mathbf{x}' \cdot \mathbf{Fl}) - l_y l_w \\ l_x^2 + l_y^2 \end{pmatrix}$$

with $\mathbf{F} = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \end{bmatrix}$.

The points $\hat{\mathbf{x}}$ and $\hat{\mathbf{x}}'$ obtained above are used to provide an initial solution in the general non-isotropic covariance case, for which closed-form solution does not exist. In the general case the non-isotropic covariance matrices $\Lambda_{\mathbf{x}}$ and $\Lambda_{\mathbf{x}'}$ are approximated with isotropic ones with radius

$$r = |det(\Lambda_{\mathbf{x}})|^{1/4} \quad r' = |det(\Lambda_{\mathbf{x}'})|^{1/4}$$

then (29) is applied and the solution end points are refined by using a Levenberg-Marquardt numerical algorithm to minimize the (24) while satisfying the alignment constraint (25).

Appendix C
Variance of distance between planes

Covariance of MLE end points

In Appendix B we have shown how to estimate the MLE points $\hat{\mathbf{x}}$ and $\hat{\mathbf{x}}'$. We here demonstrate how to compute the 4×4 covariance matrix of the MLE 4-vector $\hat{\zeta} = (\hat{\mathbf{x}}^\top \hat{\mathbf{x}}'^\top)^\top$ from the covariances of the input points $\mathbf{x}$ and $\mathbf{x}'$ and the covariance of the projection matrix.

In order to simplify the following development we define the points: $\mathbf{b} = \mathbf{x}$ on the plane π; and $\mathbf{t} = \mathbf{x}'$ on the plane π' corresponding to $\mathbf{x}$.

It can be shown that the 4×4 covariance matrix $\Lambda_{\hat{\zeta}}$ of the vector $\hat{\zeta} = \begin{pmatrix} \hat{b}_x & \hat{b}_y & \hat{t}_x & \hat{t}_y \end{pmatrix}^\top$ (MLE top and base points, see section (4.2.1)) can be computed by using the *implicit function theorem* [Cla98, Fau93] as:

$$\Lambda_{\hat{\zeta}} = \mathbf{A}^{-1}\mathbf{B}\Lambda_{\zeta}\mathbf{B}^\top\mathbf{A}^{-\top} \quad (30)$$

where $\zeta = (b_x, b_y, t_x, t_y, p_{13}, p_{23}, p_{33})^\top$ and

$$\Lambda_{\zeta} = \begin{bmatrix} \Lambda_{\mathbf{b}} & 0 & 0 \\ 0 & \Lambda_{\mathbf{t}} & 0 \\ 0 & 0 & \Lambda_{\mathbf{p}_3} \end{bmatrix} \quad (31)$$

$\Lambda_{\mathbf{b}}$ and $\Lambda_{\mathbf{t}}$ are the 2×2 covariance matrices of the points $\mathbf{b}$ and $\mathbf{t}$ respectively and $\Lambda_{\mathbf{p}_3}$ is the 3×3 covariance matrix of the vector $\mathbf{p}_3 = \alpha\mathbf{v}$ defined in (4). Note that the assumption of statistical independence in (31) is a valid one.

The matrix $\mathtt{A}$ in (30) is the the following 4×4 matrix

$$\mathtt{A} = [\, \mathtt{A}_1 \ \mathtt{A}_2 \,]$$

$$\mathtt{A}_1 = \begin{bmatrix} -\mathbf{e}_1^{\mathbf{b}} \cdot \boldsymbol{\delta}_t & -\mathbf{e}_2^{\mathbf{b}} \cdot \boldsymbol{\delta}_t \\ \delta_{e_x}\delta_{b_y} & \delta_{e_y}\delta_{b_y} - \tau\lambda p_{33} \\ \tau\lambda p_{33} - \delta_{e_x}\delta_{b_x} & -\delta_{e_y}\delta_{b_x} \\ -\tau\delta_{t_y} & \tau\delta_{t_x} \end{bmatrix}$$

$$\mathtt{A}_2 = \begin{bmatrix} -\lambda p_{33}\delta_{t_y} & \lambda p_{33}\delta_{t_x} \\ -\tau e_{11}^t - \lambda p_{33}\delta_{b_y} & -\tau e_{12}^t - \lambda p_{33}\delta_{b_y} \\ -\tau e_{12}^t + \lambda p_{33}\delta_{b_x} & -\tau e_{22}^t + \lambda p_{33}\delta_{b_x} \\ \tau\delta_{b_y} & -\tau\delta_{b_x} \end{bmatrix}$$

where we have defined:

- $\mathtt{E_t} = \Lambda_{\mathbf{t}}^{-1}$ and e_{ij}^t its ij^{th} element;
- $\mathtt{E_b} = \Lambda_{\mathbf{b}}^{-1}$ and $\mathbf{e}_1^{\mathbf{b}}$ and $\mathbf{e}_2^{\mathbf{b}}$ respectively its first and second row;
- $\mathbf{p} = (p_{13}, p_{23})^\top$, $\boldsymbol{\delta}_{\mathbf{t}} = p_{33}\hat{\mathbf{t}} - \mathbf{p}$, $\boldsymbol{\delta}_{\mathbf{b}} = p_{33}\hat{\mathbf{b}} - \mathbf{p}$, $\boldsymbol{\delta}_{\mathbf{e}} = \mathbf{e}_2^{\mathbf{b}} - \mathbf{e}_1^{\mathbf{b}}$;
- $\tau = (\mathbf{p}_3 \times \hat{\mathbf{t}})_y - (\mathbf{p}_3 \times \hat{\mathbf{t}})_x$, $\lambda = \frac{\boldsymbol{\delta}_{\mathbf{e}} \cdot (\mathbf{b} - \hat{\mathbf{b}})}{\tau}$;

The matrix $\mathtt{B}$ in (30) is the following 4×7 matrix:

$$\mathtt{B} = [\, \mathtt{B}_1 \ \mathtt{B}_2 \,]$$

$$\mathtt{B}_1 = \begin{bmatrix} \mathbf{e}_1^{\mathbf{b}} \cdot \boldsymbol{\delta}_t & \mathbf{e}_2^{\mathbf{b}} \cdot \boldsymbol{\delta}_t & 0 & 0 \\ -\delta_{e_x}\delta_{b_y} & -\delta_{e_y}\delta_{b_y} & \tau e_{11}^t & \tau e_{12}^t \\ \delta_{e_x}\delta_{b_x} & \delta_{e_y}\delta_{b_x} & \tau e_{12}^t & \tau e_{22}^t \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathtt{B}_2 = \begin{bmatrix} \lambda\delta_{t_y} & -\lambda\delta_{t_x} & -\lambda\nu_1 \\ -\lambda\delta_{b_y} & -\lambda(\tau + \delta_{b_y}) & \lambda\nu_2 \\ \lambda(\tau + \delta_{b_x}) & \lambda\delta_{b_x} & -\lambda\nu_3 \\ \tau(\hat{t}_y - \hat{b}_y) & \tau(\hat{b}_x - \hat{t}_x) & \tau\nu_4 \end{bmatrix}$$

where we have defined

$$\begin{aligned} \nu_1 &= \hat{t}_y(p_{23}\hat{t}_x - p_{13}\hat{t}_y) \\ \nu_2 &= \hat{b}_y(p_{13} + p_{23}) - p_{23}(\hat{t}_x + \hat{t}_y) \\ \nu_3 &= \hat{b}_x(p_{13} + p_{23}) - p_{13}(\hat{t}_x + \hat{t}_y) \\ \nu_4 &= \hat{t}_x\hat{b}_y - \hat{t}_y\hat{b}_x \end{aligned}$$

Note that if the vanishing point is noise-free then $\Lambda_{\hat{\zeta}}$ has rank 3 as expected because of the alignment constraint.

Variance of the distance measurement, σ_Z^2

As seen in section 4.2.1 and 4.2.2 the components of the $\hat{\boldsymbol{\zeta}}$ vector are used to compute the distance Z according to equation (9) rewritten here as:

$$Z = -\frac{||\hat{\mathbf{b}} \times \hat{\mathbf{t}}||}{(\mathbf{p}_4 \cdot \hat{\mathbf{b}})||\mathbf{p}_3 \times \hat{\mathbf{t}}||}$$

with the MLE points $\hat{\mathbf{b}}$, $\hat{\mathbf{t}}$ homogeneous with unit third coordinate.

Let us define

$$\beta = ||\hat{\mathbf{b}} \times \hat{\mathbf{t}}||, \quad \gamma = ||\mathbf{p}_3 \times \hat{\mathbf{t}}||, \quad \rho = \mathbf{p}_4 \cdot \hat{\mathbf{b}}$$

The variance σ_Z^2 of the measurement Z depends on the covariance of the $\hat{\boldsymbol{\zeta}}$ vector and the covariance of the 6-vector $\mathbf{p} = \left(\, \mathbf{p}_3^\top \ \mathbf{p}_4^\top \,\right)^\top$ computed in section 4.1. If $\hat{\boldsymbol{\zeta}}$ and $\mathbf{p}$ are statistically independent, then from first order error analysis

$$\sigma_Z^2 = \boldsymbol{\nabla}_Z \begin{pmatrix} \Lambda_{\hat{\zeta}} & 0 \\ 0 & \Lambda_{\mathbf{p}} \end{pmatrix} \boldsymbol{\nabla}_Z^\top \qquad (32)$$

the 1×10 Jacobian $\boldsymbol{\nabla}_Z$ is:

$$\boldsymbol{\nabla}_Z = Z \begin{pmatrix} \mathtt{F}\left(\frac{(\hat{\mathbf{t}} \times \hat{\mathbf{b}}) \times \hat{\mathbf{t}}}{\beta^2} - \frac{\mathbf{p}_4}{\rho}\right) \\ \mathtt{F}\left(\frac{(\hat{\mathbf{b}} \times \hat{\mathbf{t}}) \times \hat{\mathbf{b}}}{\beta^2} - \frac{(\mathbf{p}_3 \times \hat{\mathbf{t}}) \times \mathbf{p}_3}{\gamma^2}\right) \\ \frac{(\mathbf{p}_3 \times \hat{\mathbf{t}}) \times \hat{\mathbf{t}}}{\gamma^2} \\ -\frac{\hat{\mathbf{b}}}{\rho} \end{pmatrix}^\top$$

where $\mathtt{F} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$.

Note that the assumption of statistical independence in (32) is an approximation.

Appendix D
Variance of the affine parameter α

In section 8 the affine parameter α is obtained by computing the eigenvector $\mathbf{s}$ with smallest eigenvalue of the matrix $\mathtt{A}^\top\mathtt{A}$ (9). If the measured reference points are noise-free, or $n = 1$, then $\mathbf{s} = Null(\mathtt{A})$ and in general we can assume that for $\mathbf{s}$ the residual error $\mathbf{s}^\top\mathtt{A}^\top\mathtt{A}\mathbf{s} = \lambda \approx 0$.

We now use matrix perturbation theory [GVL89, SS90, Wil65] to compute the covariance $\Lambda_{\mathbf{s}}$ of the solution vector $\mathbf{s}$ based on this zero approximation.

Note that the ith row of the matrix $\mathtt{A}$ depends on the normalized vanishing line $\mathbf{l}$, on the vanishing point $\mathbf{v}$, on the reference end points $\mathbf{b}_i, \mathbf{t}_i$ and on reference distances Z_i. Uncertainty in any of those

elements induces an uncertainty in the matrix $\mathtt{A}$ and therefore uncertainty in the final solution $\mathbf{s}$.

We now define the input vector

$$\boldsymbol{\eta} \;=\; (l_x \; l_y \; l_w \; v_x \; v_y \; v_w \; Z_1 \; t_{1_x} \; t_{1_y} \; b_{1_x} \; b_{1_y} \; \cdots$$
$$Z_n \; t_{n_x} \; t_{n_y} \; b_{n_x} \; b_{n_y})^\top$$

which contains the plane vanishing line, the vanishing point and the $5n$ components of the n references. Because of noise we have:

$$\boldsymbol{\eta} \;=\; \tilde{\boldsymbol{\eta}} + \delta\boldsymbol{\eta} =$$
$$(\tilde{l}_x \; \tilde{l}_y \; \tilde{l}_w \; \tilde{v}_x \; \tilde{v}_y \; \tilde{v}_w \; \tilde{Z}_1 \; \tilde{t}_{1_x} \; \tilde{t}_{1_y} \; \tilde{b}_{1_x} \; \tilde{b}_{1_y} \cdots$$
$$\tilde{Z}_n \; \tilde{t}_{n_x} \; \tilde{t}_{n_y} \; \tilde{b}_{n_x} \; \tilde{b}_{n_y})^\top +$$
$$(\delta l_x \; \delta l_y \; \delta l_w \; \delta v_x \; \delta v_y \; \delta v_w \; \delta Z_1 \; \delta t_{1_x} \; \delta t_{1_y} \cdots$$
$$\delta Z_n \; \delta t_{n_x} \; \delta t_{n_y} \; \delta b_{n_x} \; \delta b_{n_y})^\top$$

where the '~' indicates noiseless quantities.

We assume that the noise is gaussian with zero mean and also that different reference distances are uncorrelated. However, the rows of the $\mathtt{A}$ matrix are correlated by the presence of $\mathbf{v}$ and $\mathbf{l}$ in each of them.

The 1×2 row-vector of the design matrix $\mathtt{A}$ is

$$\mathbf{a}_i = \left(\; Z_i \rho_i \gamma_i \quad \beta_i \; \right)$$

with $i = 1 \cdots n$.

Because of the noise $\mathbf{a}_i = \tilde{\mathbf{a}}_i + \delta\mathbf{a}_i$ and

$$\delta\mathbf{a}_i = \left(\; \rho_i\gamma_i\delta Z_i + Z_i\gamma_i\delta\rho_i + Z_i\rho_i\delta\gamma_i \quad \delta\beta_i \; \right)$$

It can be shown that $\delta\rho_i$, $\delta\gamma_i$ and $\delta\beta_i$ can be computed as functions of $\delta\boldsymbol{\eta}$ and therefore, taking account of the statistical dependence of the rows of the $\mathtt{A}$ matrix, the 2×2 matrices $\mathtt{E}(\delta\mathbf{a}_i^\top \delta\mathbf{a}_j) \; \forall i, j = 1 \cdots n$ can be computed.

Furthermore if we define the matrix $\mathtt{M} = \mathtt{A}^\top \mathtt{A}$ then

$$\mathtt{M} \;=\; (\tilde{\mathtt{A}} + \delta\mathtt{A})^\top (\tilde{\mathtt{A}} + \delta\mathtt{A}) =$$
$$\tilde{\mathtt{A}}^\top\tilde{\mathtt{A}} + \delta\mathtt{A}^\top\tilde{\mathtt{A}} + \tilde{\mathtt{A}}^\top\delta\mathtt{A} + \delta\mathtt{A}^\top\delta\mathtt{A}$$

Thus $\mathtt{M} = \tilde{\mathtt{M}} + \delta\mathtt{M}$ and for the first order approximation we get $\delta\mathtt{M} = \delta\mathtt{A}^\top\tilde{\mathtt{A}} + \tilde{\mathtt{A}}^\top\delta\mathtt{A}$.

As noted the vector $\mathbf{s}$ is the eigenvector corresponding to the null eigenvalue of the matrix $\tilde{\mathtt{M}}$; the other eigensolution is: $\tilde{\mathtt{M}}\tilde{\mathbf{u}}_2 = \tilde{\lambda}_2\tilde{\mathbf{u}}_2$ with $\tilde{\mathbf{u}}_2$ the second eigenvector of the $\mathtt{A}^\top\mathtt{A}$ matrix and $\tilde{\lambda}_2$ the corresponding eigenvalue.

It is proved in [GVL89, SB95] that the variation of the solutions is related to the noise of the matrix

$\mathtt{M}$ as:

$$\delta\mathbf{s} = -\frac{\tilde{\mathbf{u}}_2\tilde{\mathbf{u}}_2^\top}{\tilde{\lambda}_2}\delta\mathtt{M}\tilde{\mathbf{s}}$$

but since $\delta\mathtt{M}\tilde{\mathbf{s}} = \delta\mathtt{A}^\top\tilde{\mathtt{A}}\tilde{\mathbf{s}} + \tilde{\mathtt{A}}^\top\delta\mathtt{A}\tilde{\mathbf{s}}$ and $\tilde{\mathtt{A}}\tilde{\mathbf{s}} = 0$ then

$$\delta\mathtt{M}\tilde{\mathbf{s}} = \tilde{\mathtt{A}}^\top\delta\mathtt{A}\tilde{\mathbf{s}}$$

and thus $\delta\mathbf{s} = \tilde{\mathtt{J}}\tilde{\mathtt{A}}^\top\delta\mathtt{A}\tilde{\mathbf{s}}$ where $\tilde{\mathtt{J}}$ is simply

$$\tilde{\mathtt{J}} = -\frac{\tilde{\mathbf{u}}_2\tilde{\mathbf{u}}_2^\top}{\tilde{\lambda}_2}$$

Therefore:

$$\Lambda_{\mathbf{s}} \;=\; E\left[\delta\mathbf{s}\delta\mathbf{s}^\top\right] \tag{33}$$
$$=\; \tilde{\mathtt{J}}E\left[\tilde{\mathtt{A}}^\top\delta\mathtt{A}\tilde{\mathbf{s}}\tilde{\mathbf{s}}^\top\delta\mathtt{A}^\top\tilde{\mathtt{A}}\right]\tilde{\mathtt{J}}^\top$$
$$=\; \tilde{\mathtt{J}}E\left[\sum_{i=1}^{n}\tilde{\mathbf{a}}_i^\top\,(\delta\tilde{\mathbf{a}}_i \cdot \tilde{\mathbf{s}})\sum_{j=1}^{n}\tilde{\mathbf{a}}_j(\delta\tilde{\mathbf{a}}_j \cdot \tilde{\mathbf{s}})\right]\tilde{\mathtt{J}}^\top$$
$$=\; \tilde{\mathtt{J}}E\left[\sum_{i=1}^{n}\tilde{\mathbf{a}}_i^\top\left(\sum_{j=1}^{n}\tilde{\mathbf{a}}_j\tilde{\mathbf{s}}^\top(\delta\tilde{\mathbf{a}}_i^\top\delta\tilde{\mathbf{a}}_j)\tilde{\mathbf{s}}\right)\right]\tilde{\mathtt{J}}^\top$$
$$=\; \tilde{\mathtt{J}}\left[\sum_{i=1}^{n}\tilde{\mathbf{a}}_i^\top\left(\sum_{j=1}^{n}\tilde{\mathbf{a}}_j\tilde{\mathbf{s}}^\top E(\delta\tilde{\mathbf{a}}_i^\top\delta\tilde{\mathbf{a}}_j)\tilde{\mathbf{s}}\right)\right]\tilde{\mathtt{J}}^\top$$

having used that

$$(\delta\tilde{\mathbf{a}}_i \cdot \tilde{\mathbf{s}})(\delta\tilde{\mathbf{a}}_j \cdot \tilde{\mathbf{s}}) = \tilde{\mathbf{s}}^\top(\delta\tilde{\mathbf{a}}_i^\top\delta\tilde{\mathbf{a}}_j)\tilde{\mathbf{s}}$$

Now considering that $\tilde{\mathtt{J}}$ is a symmetric matrix ($\tilde{\mathtt{J}}^\top = \tilde{\mathtt{J}}$) eq. (33) can be written as

$$\Lambda_{\mathbf{s}} = \tilde{\mathtt{J}}\tilde{\mathtt{S}}\tilde{\mathtt{J}}$$

where $\tilde{\mathtt{S}}$ is the following 2×2 matrix:

$$\tilde{\mathtt{S}} = \sum_{i=1}^{n}\tilde{\mathbf{a}}_i^\top\left(\sum_{j=1}^{n}\tilde{\mathbf{a}}_j\tilde{\mathbf{s}}^\top\mathtt{E}_{ij}\tilde{\mathbf{s}}\right)$$

with $\mathtt{E}_{ij} = E(\delta\tilde{\mathbf{a}}_i^\top\delta\tilde{\mathbf{a}}_j)$.

Note that many of the above equations require the true noise-free quantities, which in general are not available. Weng et al. [WHA89] pointed out that if one writes, for instance, $\tilde{\mathtt{A}} = \mathtt{A} - \delta\mathtt{A}$ and substitutes this in the relevant equations, the term in $\delta\mathtt{A}$ disappears in the first order expression, allowing $\tilde{\mathtt{A}}$ to be simply interchanged with $\mathtt{A}$, and so on. Therefore the 2×2 covariance matrix $\Lambda_{\mathbf{s}}$ is simply

$$\Lambda_{\mathbf{s}} = \mathtt{J}\mathtt{S}\mathtt{J} \tag{34}$$

where $J = -\frac{\mathbf{u}_2 \mathbf{u}_2^\top}{\lambda_2}$. The 2×2 matrix S is:

$$S = \sum_{i=1}^{n} \mathbf{a}_i^\top \left(\sum_{j=1}^{n} \mathbf{a}_j \mathbf{s}^\top E_{ij} \mathbf{s} \right) \qquad (35)$$

with $\mathbf{a}_i$ the ith 1×2 row-vector of the design matrix A and n the number of references.

The 2×2 covariance matrix $\Lambda_\mathbf{s}$ of the vector $\mathbf{s}$ is therefore computed.

Noise-free $\mathbf{v}$ *and* $\mathbf{l}$.

In the case $\Lambda_\mathbf{l} = 0$ and $\Lambda_\mathbf{v} = 0$ then (35) simply becomes:

$$S = \sum_{i=1}^{n} \mathbf{a}_i^\top \mathbf{a}_i \mathbf{s}^\top E_{ii} \mathbf{s} \qquad (36)$$

in fact the rows of the A matrix are all statistically independent.

Variance of α

It is easy to convert the 2×2 homogeneous covariance matrix $\Lambda_\mathbf{s}$ in (34) into inhomogeneous coordinates. In fact, since $\mathbf{s} = \begin{pmatrix} s(1) & s(2) \end{pmatrix}^\top$ and $\alpha = \frac{s(1)}{s(2)}$ for a first order error analysis the variance of the affine parameter α is

$$\sigma_\alpha^2 = \nabla \alpha \Lambda_\mathbf{s} \nabla \alpha^\top \qquad (37)$$

with the 1×2 Jacobian

$$\nabla \alpha = \frac{1}{s(2)^2} \begin{pmatrix} s(2) & -s(1) \end{pmatrix}$$

References

[Alb35]   Leon Battista Alberti. *De Pictura*. 1435. Reproduced by Laterza (1980).

[Bar83]   S. T. Barnard. Interpreting perspective images. *Artificial Intelligence*, 21(3):435–462, 1983.

[Ber87]   M. Berger. *Geometry II*. Springer-Verlag, 1987.

[Can86]   J. F. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.

[Cla98]   J. C. Clarke. Modelling uncertainty: A primer. Technical Report 2161/98, University of Oxford, Dept. Engineering Science, 1998.

[CRZ99a]   A Criminisi, I. Reid, and A Zisserman. A plane measuring device. *Image and Vision Computing*, 17(8):625–634, 1999.

[CRZ99b]   A. Criminisi, I. Reid, and A. Zisserman. Single view metrology. In *Proc. 7th International Conference on Computer Vision, Kerkyra, Greece*, pages 434–442, September 1999.

[CT90]   B. Caprile and V. Torre. Using vanishing points for camera calibration. *International Journal of Computer Vision*, pages 127–140, 1990.

[CW90]   R. T. Collins and R. S. Weiss. Vanishing point calculation as a statistical inference on the unit sphere. In *Proc. 3rd International Conference on Computer Vision, Osaka*, pages 400–403, December 1990.

[DF95]   F. Devernay and O. D. Faugeras. Automatic calibration and removal of distortion from scenes of structured environments. In *SPIE*, volume 2567, San Diego, CA, July 1995.

[Fau93]   O. D. Faugeras. *Three-Dimensional Computer Vision: a Geometric Viewpoint*. MIT Press, 1993.

[GVL89]   G. H. Golub and C.F. Van Loan. *Matrix Computations*. The John Hopkins University Press, Baltimore, MD, second edition, 1989.

[HAA97]   Y. Horry, K. Anjyo, and K. Arai. Tour into the picture: Using a spidery mesh interface to make animation from a single image. In *Proceedings of the ACM SIGGRAPH Conference on Computer Graphics*, pages 225–232, 1997.

[KSH98]   T. Kim, Y. Seo, and K. Hong. Physics-based 3D position analysis of a soccer ball from monocular image sequences. *Proc. International Conference on Computer Vision*, pages 721 – 726, 1998.

[KvD91]   J. J. Koenderink and A. J. van Doorn. Affine structure from motion. *J. Opt. Soc. Am. A*, 8(2):377–385, 1991.

[LCZ99]   D. Liebowitz, A. Criminisi, and A. Zisserman. Creating architectural models from images. In *Proc. EuroGraphics*, volume 18, pages 39–50, September 1999.

[LZ98]   D. Liebowitz and A. Zisserman. Metric rectification for perspective images of planes. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 482–488, June 1998.

[MK95]   G. F. McLean and D. Kotturi. Vanishing point detection by line clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(11):1090–1095, 1995.

[PTVG98]   M. Proesmans, T. Tuytelaars, and L. J. Van Gool. Monocular image measurements. Technical Report Improofs-M12T21/1/P, K.U.Leuven, 1998.

[QM92]   L. Quan and R. Mohr. Affine shape representation from motion through reference points. *Journal of Mathematical Imaging and Vision*, 1:145–151, 1992.

[RF93]   L. Robert and O. D. Faugeras. Relative 3D positioning and 3D convex hull computation from a weakly calibrated stereo pair. In *Proc. 4th International Conference on Computer Vision, Berlin*, pages 540–544, 1993.

[RN98]   I. D. Reid and A. North. 3D trajectories from a single viewpoint using shadows. *Proc. British Machine Vision Conference*, 1998.

[RZ96]   I. Reid and A. Zisserman. Goal-directed video metrology. In R. Cipolla and B. Buxton, editors, *Proc. 4th European Conference on Computer Vision, LNCS 1065, Cambridge*, volume II, pages 647–658. Springer, April 1996.

[SB95]   L. S. Shapiro and J. M. Brady. Rejecting outliers and estimating errors in an orthogonal regression framework. *Philosophical Transactions of the Royal Society of London, SERIES A*, 350:407–439, 1995.

[Shu99]   J. A. Shufelt. Performance and analysis of vanishing point detection techniques. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(3):282–288, March 1999.

[SM99]   P. Sturm and S. Maybank. A method for interactive 3D reconstruction of piecewise planar objects from single images. In *Proc. 10th British Machine Vision Conference, Nottingham*, 1999.

[Spr64]   C. E. Springer. *Geometry and Analysis of Projective Spaces*. Freeman, 1964.

[SS90]   G. W. Stewart and J. Sun. *Matrix Perturbation Theory*. Academic Press Inc., USA, 1990.

[VGPZ98] L. Van Gool, M. Proesmans, and A. Zisserman. Planar homologies as a basis for grouping and recognition. *Image and Vision Computing*, 16:21–26, January 1998.

[VL99]   T. Viéville and D. Lingrand. Using specific displacements to analyze motion without calibration. *International Journal of Computer Vision*, 31(1):5–30, February 1999.

[WHA89]   J. Weng, T. S. Huang, and N. Ahuja. Motion and structure from two perspective views: algorithms, error analysis and error estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(5):451–476, 1989.

[Wil65]   J. H. Wilkinson. *The Algebraic Eigenvalue Problem*. Clarendon Press, Oxford, 1965.