

# Single View Metrology

A. Criminisi, I. Reid and A. Zisserman \*

Department of Engineering Science

University of Oxford

Oxford, UK, OX1 3PJ

{criminisi,ian,az}@robots.ox.ac.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 (viewer'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.*

## 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 [8, 9, 13, 14].

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 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 ideas in this paper can be seen as reversing the rules for drawing perspective images given by Leon Battista Alberti [1] 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.

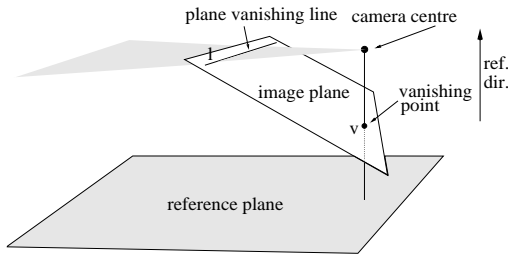
We begin in section 2 by giving geometric interpretations for the key scene features, and then give simple geometric derivations of how, in principle, three dimensional affine information may be extracted from the image. In section 3 we introduce an algebraic representation of the problem and show that 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 two important ones: forensic measurement and virtual modelling in section 5.

## 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 [3, 11, 12]. Ef-

---

\*The authors would like to thank Andrew Fitzgibbon for assistance with the TargetJr libraries, and David Liebowitz and Luc van Gool for discussions. This work was supported by the EU Esprit Project IMPROOFS.



**Figure 1: 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. The vanishing point  $v$  is the intersection of the image plane with a line parallel to the reference direction through the camera centre.

fects 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 [6], and are therefore not detrimental to our methods (see, for example, figure 9).

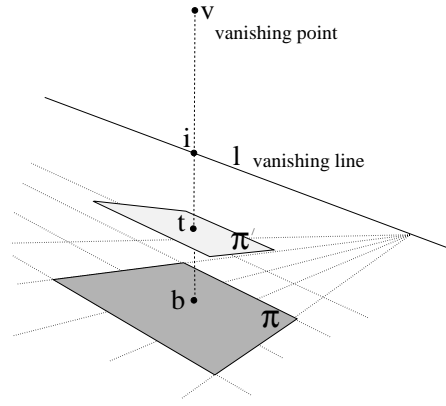
Although the schematic figures show the camera centre at a finite location, the results we derive apply also to the case of a camera centre at infinity, i.e. where the images are obtained by parallel projection. The basic geometry of the plane’s vanishing line and the vanishing point are illustrated in figure 1. 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.

It can be seen (for example, by inspection of figure 1) 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 further from the plane, and if “below” the vanishing line, then it is closer to the plane than the camera centre.

Two points on separate planes (parallel to the reference plane) *correspond* if the line joining them is parallel to the reference direction; hence the image of each point 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.

### 2.1. Measurements between parallel planes

We wish to measure the distance between two parallel planes, specified by the image points  $t$  and  $b$ , in the reference direction. Figure 2 shows the geometry, with points  $t$  and  $b$  in correspondence. The four points marked on the figure define a cross-ratio. The vanishing point is the image of a point at infinity in the scene [15]. In the image the value



**Figure 2: Cross ratio:** The point  $b$  on the plane  $\pi$  corresponds to the point  $t$  on the plane  $\pi'$ . They are aligned with the vanishing point  $v$ . The four points  $v$ ,  $t$ ,  $b$  and the intersection  $i$  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.

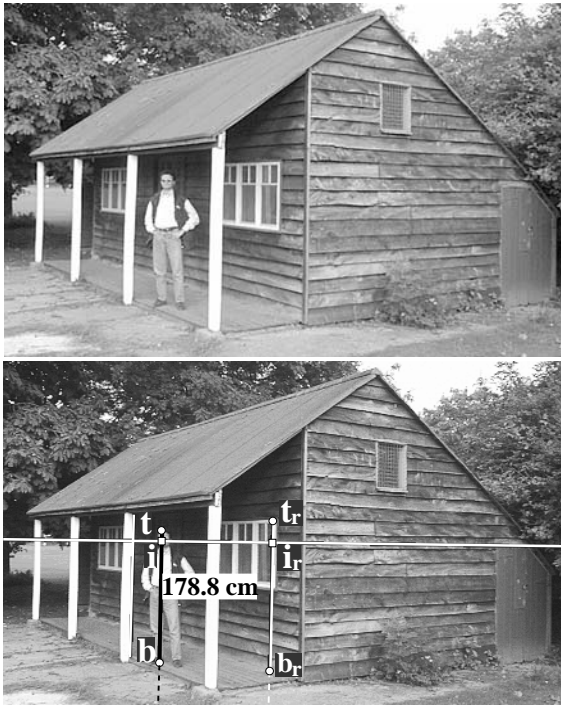
of the cross-ratio provides an affine length ratio. In fact we obtain the ratio of the distance between the planes containing  $t$  and  $b$ , to the camera’s distance from the plane  $\pi$  (or  $\pi'$  depending on the ordering of the cross-ratio). The absolute distance can be obtained from this distance ratio once the camera’s distance from  $\pi$  is specified. However it is usually more practical to determine the distance via a second measurement in the image, that of a known reference length.

Furthermore, since the vanishing line is the imaged axis of the pencil of planes parallel to the reference plane, the knowledge of the distance between *any* pair of the planes is sufficient to determine the absolute distance between another two of the planes.

**Example.** Figure 3 shows that a person’s height may be computed from an image given a vertical reference height elsewhere in the scene. The formula used to compute this result is given in section 3.1.

### 2.2. Measurements on parallel planes

If the reference plane  $\pi$  is affine calibrated (we know its vanishing line) then from image measurements we can compute: (i) ratios of lengths of parallel line segments on the plane; (ii) 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.

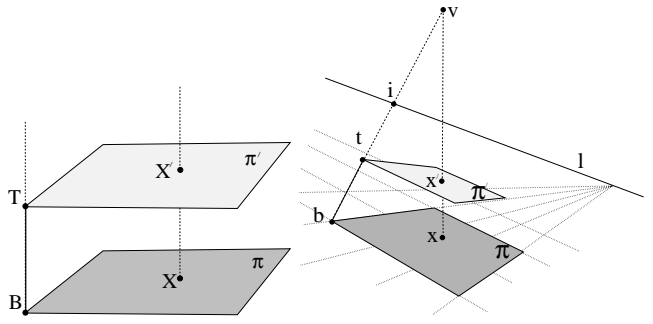


**Figure 3: Measuring the height of a person:** (top) original image; (bottom) the height of the person is computed from the image as 178.8cm (the true height is 180cm, but note that the person is leaning down a bit on his right foot). The vanishing line is shown in white and the reference height is the segment  $(t_r, b_r)$ . The vertical vanishing point is not shown since it lies well below the image.  $t$  is the top of the head and  $b$  is the base of the feet of the person while  $i$  is the intersection with the vanishing line.

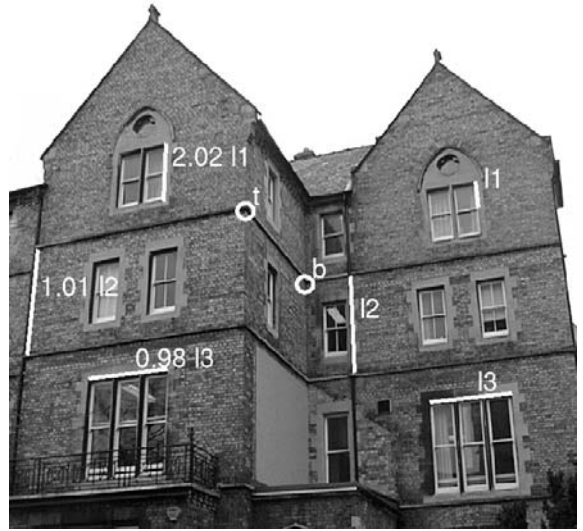
A map in the world between parallel planes induces a map in the image between images of points on the two planes. This image map is a *planar homology* [15], which is a plane projective transformation with five degrees of freedom, having a line of fixed points, called the *axis* and a distinct fixed point not on the axis known as the *vertex*. Planar homologies arise naturally in an image when two planes related by a perspectivity in 3-space are imaged [16]. The geometry is illustrated in figure 4.

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. This line and point specify four of the five degrees of freedom of the homology. The remaining degree of freedom of the homology is uniquely determined from any pair of image points which correspond between the planes (points  $b$  and  $t$  in figure 4).

This means that we can compare measurements made on two separate planes by mapping between the planes in the reference direction via the homology. In particular we may compute (i) the ratio between two parallel lengths, one length on each plane; (ii) 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 homol-



**Figure 4: Homology mapping between parallel planes:** (left) A point  $X$  on plane  $\pi$  is mapped into the point  $X'$  on  $\pi'$  by a parallel projection. (right) In the image the mapping between the images of the two planes is a homology, with  $v$  the vertex and  $l$  the axis. The correspondence  $b \rightarrow t$  fixes the remaining degree of freedom of the homology from the cross-ratio of the four points:  $v, i, t$  and  $b$ .



**Figure 5: Measuring the ratio of lengths of parallel line segments lying on two parallel scene planes:** The points  $t$  and  $b$  (together with the plane vanishing line and the vanishing point) define the homology between the two planes on the facade of the building.

ogy and then, since the reference plane's vanishing line is known, make affine measurements in the plane, e.g. parallel length or area ratios.

**Example.** Figure 5 shows that given the reference vanishing line and vanishing point, and a point correspondence (in the reference direction) on each of two parallel planes, then the ratio of lengths of parallel line segments may be computed from the image. The formula used to compute this result is given in section 3.2.

### 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 from a particular plane knowing a single reference distance.

Furthermore, by considering figure 1 it is seen that the location of the camera relative to the reference plane is the back-projection of the 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.

We show an example in figure 12, where the location of the camera centre has been determined, and superimposed into a virtual view of the scene.

### 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 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. Let the origin of the coordinate frame lie on the reference plane, 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  $3 \times 4$  projection matrix  $\mathbf{P}$  as:

$$\mathbf{x} = \mathbf{P}\mathbf{X} = [\mathbf{p}_1 \quad \mathbf{p}_2 \quad \mathbf{p}_3 \quad \mathbf{p}_4] \mathbf{X}$$

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 that the first three columns of  $\mathbf{P}$  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  $\mathbf{P}$  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 points fixes the  $X$  and  $Y$  affine coordinate axes. We denote the vanishing line by  $\mathbf{l}$ , and to emphasise 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 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  $\mathbf{o} = \mathbf{p}_4 = \mathbf{l}/\|\mathbf{l}\| = \hat{\mathbf{l}}$ .

Therefore the final parametrization of the projection matrix  $\mathbf{P}$  is:

$$\mathbf{P} = [\mathbf{l}_1^\perp \quad \mathbf{l}_2^\perp \quad \alpha \mathbf{v} \quad \hat{\mathbf{l}}] \quad (1)$$

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

In the following sections we show how to compute various measurements from this projection matrix. Measurements between planes are independent of the first two (under-determined) columns of  $\mathbf{P}$ . For these measurements the only unknown quantity is  $\alpha$ . Coordinate measurements within the planes depend on the first two and the fourth columns of  $\mathbf{P}$ . They 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  $\mathbf{P}$ . 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

We wish to measure the distance between scene planes specified by a base point  $\mathbf{B}$  on the reference plane and top point  $\mathbf{T}$  in the scene. These points may be chosen as respectively  $(X, Y, 0)$  and  $(X, Y, Z)$ , and their images are  $\mathbf{b}$  and  $\mathbf{t}$ . If  $\mathbf{P}$  is the projection matrix then the image coordinates are

$$\mathbf{b} = \mathbf{P} \begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{t} = \mathbf{P} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

The equations above can be rewritten as

$$\mathbf{b} = \rho(X\mathbf{p}_1 + Y\mathbf{p}_2 + \mathbf{p}_4) \quad (2)$$

$$\mathbf{t} = \mu(X\mathbf{p}_1 + Y\mathbf{p}_2 + Z\mathbf{p}_3 + \mathbf{p}_4) \quad (3)$$

where  $\rho$  and  $\mu$  are unknown scale factors, and  $\mathbf{p}_i$  is the  $i$ th column of the  $\mathbf{P}$  matrix.

Taking the scalar product of (2) with  $\hat{\mathbf{l}}$  yields  $\rho = \hat{\mathbf{l}} \cdot \mathbf{b}$ , and combining this with the third column of (1) and (3) we obtain

$$\alpha Z = \frac{-\|\mathbf{b} \times \mathbf{t}\|}{(\hat{\mathbf{l}} \cdot \mathbf{b})\|\mathbf{v} \times \mathbf{t}\|} \quad (4)$$

Since  $\alpha Z$  scales linearly with  $\alpha$  we have obtained affine structure. If  $\alpha$  is known, then we immediately obtain a metric value for  $Z$ . Conversely, if  $Z$  is known (i.e. it is a reference distance) then we have a means of computing  $\alpha$ , and hence removing the affine ambiguity.



**Figure 6: Measuring heights using parallel lines:** Given the vertical vanishing point, the vanishing line for the ground plane and a reference height, the distance of the top of the window on the right wall from the ground plane is measured from the distance between the two horizontal lines shown, one defined by the top edge of the window, and the other on the ground plane.

**Example.** In figure 6 heights from the ground plane are measured between two parallel lines, one off the plane (top) and one on the plane (base). In fact, thanks to the plane vanishing line, given one line parallel to the reference plane it is easy to compute the family of parallel lines. Computing the distance between them is a straightforward application of (4).

### 3.2. Measurements on parallel planes

The projection matrix  $P$  from the world to the image is defined above with respect to a coordinate frame on the reference plane. In this section we determine the projection matrix  $P'$  referred to the parallel plane  $\pi'$  and we show how the homology between the two planes can be derived directly from the two projection matrices.

Suppose the world coordinate system is translated from the plane  $\pi$  onto the plane  $\pi'$  along the reference direction, then it is easy to show that we can parametrize the new projection matrix  $P'$  as:

$$P' = [p_1 \quad p_2 \quad \alpha v \quad \alpha Z v + \hat{l}]$$

where  $Z$  is the distance between the planes. Note that if  $Z = 0$  then  $P' = P$  correctly.

The plane to image homographies can be extracted from the projection matrices ignoring the third column, to give:

$$H = [p_1 \quad p_2 \quad \hat{l}], \quad H' = [p_1 \quad p_2 \quad \alpha Z v + \hat{l}]$$

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

A short computation gives the homology matrix  $\tilde{H}$  as:

$$\tilde{H} = I + \alpha Z v \hat{l}^T \quad (5)$$

Given the homology between two planes in the pencil we can transfer all points from one plane to the other and make affine measurements in the plane (see fig 5 and fig 7).



**Figure 7: Measuring ratios of areas on separate planes:** The image points  $t$  and  $b$  together with the vanishing line of the two parallel planes and the vanishing point for the orthogonal direction define the homology between the planes. The ratio between the area of the window on the left plane and that of the window on the right plane is computed.

### 3.3. Determining camera position

Suppose the camera centre is  $C = (X_c, Y_c, Z_c, W_c)^T$  in affine coordinates (see figure 1). Then since  $PC = 0$  we have

$$PC = l_1^\perp X_c + l_2^\perp Y_c + \alpha v Z_c + \hat{l} W_c = 0 \quad (6)$$

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

$$\begin{aligned} X_c &= -\det[l_2^\perp \quad v \quad \hat{l}], \quad Y_c = \det[l_1^\perp \quad v \quad \hat{l}] \\ \alpha Z_c &= -\det[l_1^\perp \quad l_2^\perp \quad \hat{l}], \quad W_c = \det[l_1^\perp \quad l_2^\perp \quad v] \end{aligned} \quad (7)$$

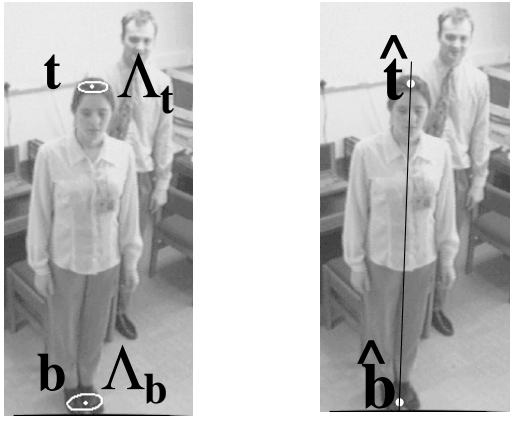
Note that once again we obtain structure off the plane up to the affine scale factor  $\alpha$ . As before, we may upgrade the distance to metric with knowledge of  $\alpha$ , or use knowledge of camera height to compute  $\alpha$  and upgrade the affine structure.

Note that affine viewing conditions (where the camera centre is at infinity) present no problem to the expressions in (7), since in this case we have  $\hat{l} = [0 \quad 0 \quad *]^T$  and  $v = [* \quad * \quad 0]^T$ . Hence  $W_c = 0$  so we obtain a camera centre on the plane at infinity, as we would expect. This point on  $\pi_\infty$  represents the viewing direction for the parallel projection.

If the viewpoint is finite (i.e. not affine viewing conditions) then the formula for  $\alpha Z_c$  may be developed further by taking the scalar product of both sides of (6) with the vanishing line  $\hat{l}$ . The result is:  $\alpha Z_c = -(\hat{l} \cdot v)^{-1}$ .

## 4. Uncertainty Analysis

Feature detection and extraction – whether manual or automatic (e.g. using an edge detector) – can only be achieved



**Figure 8: Maximum likelihood estimation of the top and base points (closeup of fig. 9):** (left) The top and base uncertainty ellipses, respectively  $\Lambda_t$  and  $\Lambda_b$ , are shown. These ellipses are specified by the user, and indicate a confidence region for localizing the points. (right) MLE top and base points  $\hat{t}$  and  $\hat{b}$  are aligned with the vertical vanishing point (outside the image).

to a finite accuracy. Any features extracted from an image, therefore, are subject to errors. In this section we consider how these errors propagate through the measurement formulae in order to quantify the uncertainty on the final measurements.

When making measurements between planes, uncertainty arises from the uncertainty in  $P$ , and from the uncertain image locations of the top and base points  $t$  and  $b$ . The uncertainty in  $P$  depends on the location of the vanishing line, the location of the vanishing point, and on  $\alpha$ , the affine scale factor. Since only the final two columns contribute, we model the uncertainty in  $P$  as a  $6 \times 6$  homogeneous covariance matrix,  $\Lambda_P$ . Since the two columns have only five degrees of freedom (two for  $v$ , two for  $l$  and one for  $\alpha$ ), the covariance matrix is singular, with rank five. Details of its computation are given in [4] and are omitted here for brevity.

Likewise, the uncertainty in the top and base points (resulting largely from the finite accuracy with which these features may be located in the image) is modelled by covariance matrices  $\Lambda_b$  and  $\Lambda_t$ . Since in the error-free case, these points must be aligned with the vertical vanishing point we can determine maximum likelihood estimates of their true locations ( $\hat{t}$  and  $\hat{b}$ ) by minimising the objective

$$(\mathbf{b}_2 - \hat{\mathbf{b}}_2)^\top \Lambda_{b_2}^{-1} (\mathbf{b}_2 - \hat{\mathbf{b}}_2) + (\mathbf{t}_2 - \hat{\mathbf{t}}_2)^\top \Lambda_{t_2}^{-1} (\mathbf{t}_2 - \hat{\mathbf{t}}_2)$$

(which is the sum of the Mahalanobis distances between the input points and the ML estimates, the subscript 2 indicates inhomogeneous 2-vectors) *subject to the alignment constraint*  $v \cdot (\hat{t} \times \hat{b}) = 0$ . Using standard techniques [7] we obtain a first order approximation to the  $4 \times 4$ , rank three covariance of the parameters  $\hat{z} = (\hat{t}_2^\top \hat{b}_2^\top)^\top$ . Figure 8 illustrates the idea.



**Figure 9: Uncertainty analysis on height measurements:** The image shown was captured from a cheap security type camera which exhibited radial distortion. This has been corrected and the height of the man estimated (measurements are in cm). (left) The height of the man and the associated uncertainty are computed as 190.6cm (c.f. ground truth value 190cm). The vanishing line for the ground plane is shown in white at the top of the image. When one reference height is used the uncertainty (3-sigma) is  $\pm 4.1$ cm, while (right) it reduces to  $\pm 2.9$ cm as two more reference heights are introduced (the filing cabinet and the table on the left).

Now, assuming the statistical independence of  $\hat{z}$  and  $P$  we obtain a first order approximation for the variance of the distance measurement:

$$\sigma_h^2 = \nabla_h \begin{pmatrix} \Lambda_z & 0 \\ 0 & \Lambda_P \end{pmatrix} \nabla_h^\top \quad (8)$$

where  $\nabla_h$  is the  $1 \times 10$  Jacobian matrix of the function which maps the projection matrix and top and base points to a distance between them (4). The validity of all approximation has been tested by Monte Carlo simulations and by a number of measurements on real images where the ground truth was known.

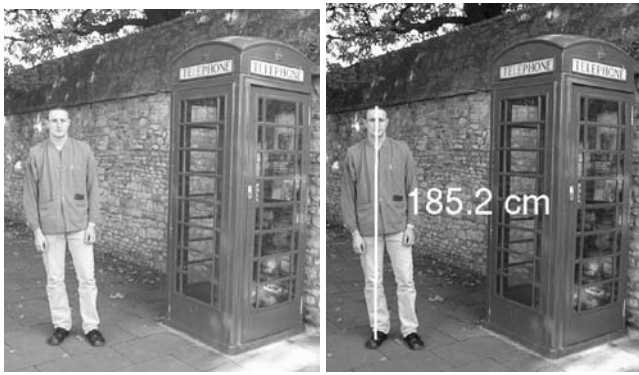
**Example.** An image obtained from a poor quality security camera is shown in figure 9. It has been corrected for radial distortion using the method described in [6], and the floor taken as the reference plane. Vertical and horizontal lines are used to compute the  $P$  matrix of the scene. One reference height is used to obtain the affine scale factor  $\alpha$  from (4), so other measurements in the same direction are metric.

The computed height of the man and an associated 3-standard deviation uncertainty are displayed in the figure. The height obtained differs by only 6mm from the known true value. As the number of reference distances is increased, so the uncertainty on  $P$  (in fact just on  $\alpha$ ) decreases, resulting in a decrease in uncertainty of the measured height, as theoretically expected.

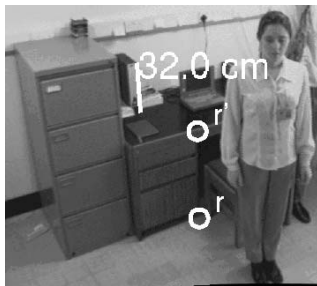
## 5. Applications

### 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.



**Figure 10: Measuring the height of a person in an outdoor scene:** The ground plane is the reference plane, and its vanishing line is computed from the slabs on the floor. The vertical vanishing point is computed from the edges of the phone box, whose height is known and used as reference. The veridical height is 187cm, but note that the person is leaning slightly on his right foot.



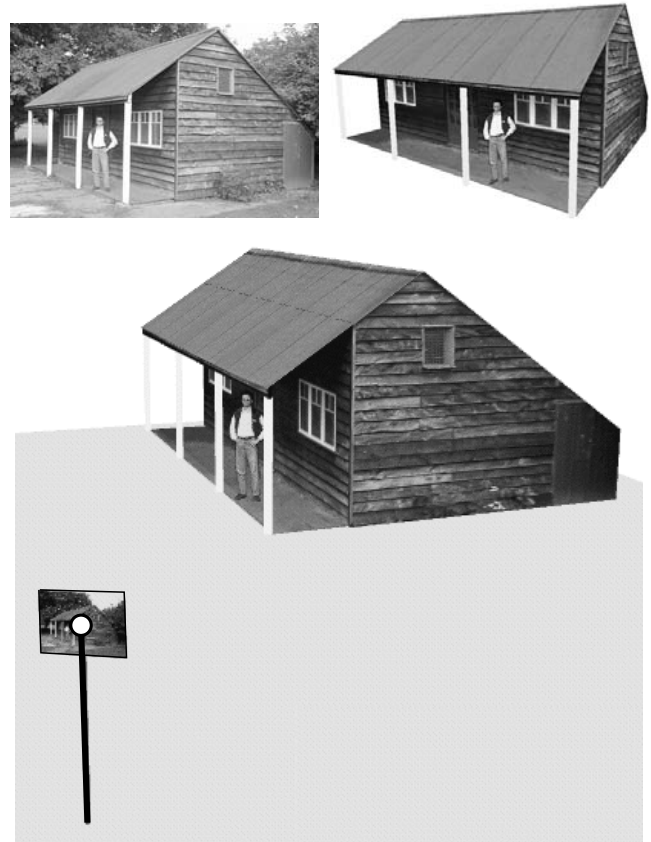
**Figure 11: Measuring heights of objects on separate planes:** Using the homology between the ground plane (initial reference) and the plane of the table, we can determine the height of the file on the table.

In figure 10 the edges of the paving stones on the floor are used to compute the vanishing line of the ground plane; the edges of the phonebox to compute the vertical vanishing point; and the height of the phone box provides the metric calibration in the vertical direction. The height of the person is then computed using (4).

Figure 11 shows an example where the homology is used to project points between planes so that a vertical distance may be measured given the distance between a plane and the reference plane.

## 5.2. Virtual modelling

In figure 12 we show an example of complete 3D reconstruction of a scene. 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. Four points with known Euclidean coordinates determine the metric calibration of the ground plane and thus for the pencil of horizontal planes which share the vanishing line. The distance of the top of the window to the ground, and the height of one of the pillars are used as reference



**Figure 12: Complete 3D reconstruction of a real scene:** (left) original image; (right) a view of the reconstructed 3D model; (bottom) 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.

lengths. The position of the camera centre is also estimated and is shown in the figure.

## 5.3. Modelling paintings

Figure 13 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 we can apply the methods developed here to obtain a correct 3D reconstruction of the scene.

Unlike other techniques [8] 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.

In the painting analyzed here, the ground plane is chosen as reference and its vanishing line can be 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. Furthermore the ground plane can be rectified from the square floor patterns and therefore the position on the ground of each

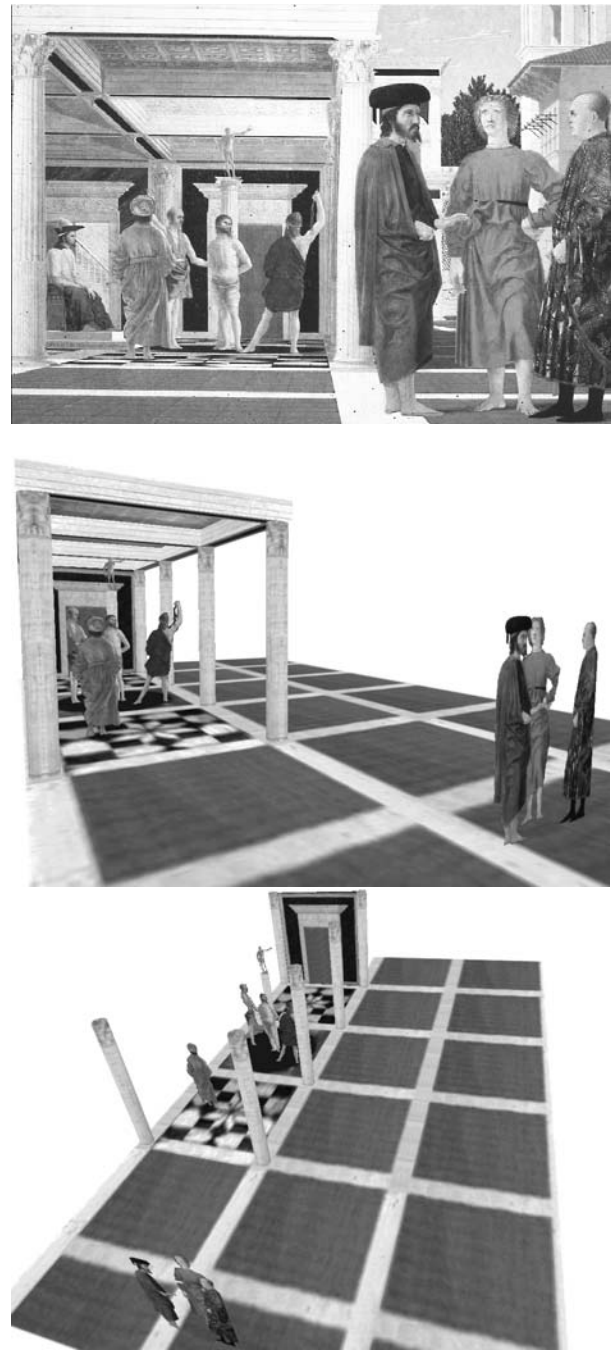
vertical object estimated [5, 10]. The measurements, up to an overall scale factor, are used to compute a three dimensional VRML model of the scene. Two different views of the model are shown in figure 13.

## 6. Summary and Conclusions

We have explored how the affine structure of 3-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. More generally, affine 3 space may be represented entirely by sets of parallel planes and directions [2]. We are currently investigating how this full geometry is best represented and computed from a single perspective image.

## References

- [1] L. B. Alberti. *De Pictura*. Laterza, 1980.
- [2] M. Berger. *Geometry II*. Springer-Verlag, 1987.
- [3] R. T. Collins and R. S. Weiss. Vanishing point calculation as a statistical inference on the unit sphere. In *Proc. ICCV*, pages 400–403, Dec 1990.
- [4] A. Criminisi, I. Reid, and A. Zisserman. Computing 3D euclidean distance from a single view. Technical Report OUEL 2158/98, Dept. Eng. Science, University of Oxford, 1998.
- [5] A. Criminisi, I. Reid, and A. Zisserman. A plane measuring device. *Image and Vision Computing*, 17(8):625–634, 1999.
- [6] F. Devernay and O. Faugeras. Automatic calibration and removal of distortion from scenes of structured environments. In *SPIE*, volume 2567, San Diego, CA, Jul 1995.
- [7] O. Faugeras. *Three-Dimensional Computer Vision: a Geometric Viewpoint*. MIT Press, 1993.
- [8] Y. Horry, K. Anjyo, and K. Arai. Tour into the picture: Using a spidery mesh interface to make animation from a single image. In *Proc. ACM SIGGRAPH*, pages 225–232, 1997.
- [9] T. Kim, Y. Seo, and K. Hong. Physics-based 3D position analysis of a soccer ball from monocular image sequences. *Proc. ICCV*, pages 721 – 726, 1998.
- [10] D. Liebowitz, A. Criminisi, and A. Zisserman. Creating architectural models from images. In *Proc. EuroGraphics*, Sep 1999.
- [11] D. Liebowitz and A. Zisserman. Metric rectification for perspective images of planes. In *Proc. CVPR*, pages 482–488, Jun 1998.
- [12] G. F. McLean and D. Kotturi. Vanishing point detection by line clustering. *IEEE T-PAMI*, 17(11):1090–1095, 1995.
- [13] M. Proesmans, T. Tuytelaars, and L. Van Gool. Monocular image measurements. Technical Report Improofs-M12T21/1/P, K.U.Leuven, 1998.
- [14] I. Reid and A. Zisserman. Goal-directed video metrology. In R. Cipolla and B. Buxton, editors, *Proc. ECCV*, volume II, pages 647–658. Springer, Apr 1996.
- [15] C. E. Springer. *Geometry and Analysis of Projective Spaces*. Freeman, 1964.
- [16] L. Van Gool, M. Proesmans, and A. Zisserman. Planar homologies as a basis for grouping and recognition. *Image and Vision Computing*, 16:21–26, Jan 1998.



**Figure 13: Complete 3D reconstruction of a Renaissance painting:** (top) *La Flagellazione di Cristo*, (1460, Urbino, Galleria Nazionale delle Marche). (middle) 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. (bottom) 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.