EUROGRAPHICS '99 / P. Brunet and R. Scopigno
(Guest Editors)

Volume 18 (*1999*), *Number 3*

Creating Architectural Models from Images

David Liebowitz, Antonio Criminisi and Andrew Zisserman

Visual Geometry Group, Department of Engineering Science, University of Oxford, Oxford, UK
{dl,criminis,az}@robots.ox.ac.uk

Abstract

We present methods for creating 3D graphical models of scenes from a limited numbers of images, i.e. one or two, in situations where no scene co-ordinate measurements are available. The methods employ constraints available from geometric relationships that are common in architectural scenes — such as parallelism and orthogonality — together with constraints available from the camera. In particular, by using the circular points of a plane simple, linear algorithms are given for computing plane rectification, plane orientation and camera calibration from a single image. Examples of image based 3D modelling are given for both single images and image pairs.

1. Introduction

The task of reconstructing objects such as buildings from photographs is receiving increased attention in the effort to create models of valuable architectural sites. We are addressing here the key case where it is not possible to make measurements of features of a scene to allow reconstruction — for example cases where buildings are destroyed and only archive images are available. In the absence of direct measurement we wish to exploit geometric characteristics such as the parallelism and orthogonality of lines and planes. Such relationships are plentiful in manmade structures, and often provide sufficient information to produce realistic reconstructions.

The techniques presented here are aimed at *metric reconstruction*; correct representation of angles and length ratios, but not of absolute scale. This level of reconstruction is precisely that required for a graphical 3D model where the absolute pose (rotation and translation) and scale are not necessary for visualization. Computation of global scale provides no theoretical difficulty, but requires knowledge of a single length measurement in the scene.

Part of the novelty of the work lies in the direct application of ideas from projective geometry[21], such as the circular points. It will be seen that this allows simple linear equations to be formulated, and avoids the non-linear constraints which typically arise in these types of applications when the orthogonality properties of rotation matrices are used. It also enables constraints from the scene and camera to be combined effortlessly when computing reconstructions, allowing the efficient use of all the available information.

We present methods to: metric rectify individual planes, compute relative perpendicular distances from partially rectified planes, calibrate cameras and reconstruct piecewise planar objects from a *single* view. We also present a method of metric rectifying 3D reconstructions from two views.

These techniques allow architectural models to be built and rendered from single images in a similar manner to the photogrammetric techniques of Debevec *et al.*[7]. However, in the Debevec system multiple images are necessary, scene measurements are required to position the cameras, and the camera internal calibration must be known. These are not needed here. We use several vanishing points for scene modelling from single images, and are therefore extending the work of Horry *et al.*[15], where a single vanishing point is used. The single view techniques are complementary to reconstruction methods applicable to multiple images, such as an image sequence acquired by a video camera when walking around a building[1, 24, 26].

In section 2 we begin with a description of planar rectification from single view scene constraints. Section 3 describes 3D modelling when one plane is partially rectified. This is followed by methods of camera calibration, in section 4. Calibration is relevant in its own right and also as an aid to the single view reconstruction of structures such as buildings, which is detailed in section 5. We briefly discuss metric reconstruction from two views using vanishing

points in section 6. Finally, some implementation details are presented in section 7.

2. Plane rectification

Given a perspective image of a world plane the goal is to obtain a metric rectification of the plane. This is equivalent to obtaining an image of the world plane where the camera's image plane and world plane are parallel.

One way to proceed is to determine the orientation of the world plane. This requires some knowledge of the internal parameters of the camera, which we return to in section 4.5. The method described in this section requires no knowledge of the camera and proceeds directly from geometric relationships on the world plane, such as parallel lines.

As is well known the map between a world plane and a perspective image is a homography (plane projective transformation)[12, 21, 23]. This map can be determined from the correspondence of four (or more) points with known position. Once the homography is determined the image can be warped onto the world plane and in this way a metric rectification is obtained. However, it is not necessary to determine the entire homography in order to obtain a metric rectification — the plane rotation, translation and uniform scaling which are a part of the homography map, and account for four degrees of freedom, are irrelevant to the rectification. This is the idea that is developed in this section: it is only necessary to determine four of the eight parameters of the homography. These four parameters are associated in projective geometry with the position of two points known as the *circular points*[21].

To introduce some notation. A homography is represented by a 3×3 homogeneous matrix $\mathtt{H}$. Points on the image plane, $\mathbf{x}$, are mapped to points on the world plane, $\mathbf{x}'$, as $\mathbf{x}' = \mathtt{H}\mathbf{x}$, where $\mathbf{x}$ is a homogeneous column 3-vector $\mathbf{x} = (x, y, 1)^\top$ with (x, y) the Euclidean position on the plane. Note, for equations between homogeneous quantities '=' is equality up to an overall scale factor. The homography matrix has eight degrees of freedom — there are nine matrix elements, but the overall scale is not significant. A line $\mathbf{l}$ is also represented by a homogeneous column 3-vector, such that if a point $\mathbf{x}$ lies on $\mathbf{l}$ then $\mathbf{l}^\top \mathbf{x} = 0$.

2.1. Metric rectification

A homography can be decomposed into two transformations:

$$\mathtt{H} = \mathtt{M}\,\mathtt{N} \tag{1}$$

The transformation $\mathtt{M}$ is the metric part of the homography and is a similarity transformation

$$\mathtt{M} = \begin{pmatrix} s\mathtt{R} & \mathbf{t} \\ \mathbf{0}^\top & 1 \end{pmatrix} \tag{2}$$

where $\mathtt{R}$ is a rotation matrix, $\mathbf{t}$ a translation vector, and s an isotropic scaling. There are four degrees of freedom in $\mathtt{M}$.

The second component of $\mathtt{H}$ is the non-metric part, which may be parametrized as

$$\mathtt{N} = \begin{pmatrix} \frac{1}{\beta} & -\frac{\alpha}{\beta} & 0 \\ 0 & 1 & 0 \\ l_1 & l_2 & 1 \end{pmatrix} \tag{3}$$

This matrix also has four degrees of freedom.

We can ignore the metric part (rotation, translation and uniform scaling) in rectifying a plane. The key point is that rectification is thus reduced to a *four parameter* problem since we only require $\mathtt{N}$ to metric rectify an image.

The transformation $\mathtt{N}$ is determined directly from projection of the circular points. The circular points on the world plane are a complex conjugate point pair with co-ordinates $(1,\ \pm i,\ 0)^\top$ lying on the line at infinity. Under the homography from the world plane to the image, $\mathtt{H}^{-1}$, the circular points are imaged as

$$\mathbf{I} = \mathtt{H}^{-1}(1,\ i,\ 0)^\top = (\alpha - i\beta, 1, -l_2 - \alpha l_1 + i l_1 \beta)^\top \tag{4}$$

and $\mathbf{J} = conj(\mathbf{I})$. Clearly, the image of the circular points depends only on the non-metric component, $\mathtt{N}$, of the homography. Furthermore, once the imaged circular points, $\mathbf{I}, \mathbf{J}$, are identified in the image, the parameters α, β, l_1, l_2 are known so that $\mathtt{N}$ may be computed.

Figure 1: *An image with significant perspective distortion. Four points, corresponding to the corners of the rectangular window, fully define a rectification homography for the plane of the building facade.*

Consider, for example, figure 1. Suppose we know the relative positions of the four points shown on the building facade and we can measure their position in the image. From these four correspondences we compute the homography relating the facade plane to its image. We thus rectify the plane as shown in figure 2. An alternative rectification method is to project the circular points into the image using (4), and obtain $\mathtt{N}$ from (3). Rectification of the plane by $\mathtt{N}$ alone gives the result shown in figure 3. The metric information on the plane is the same in both rectified images. They differ only in

Figure 2: *A metric rectified image of the plane.*

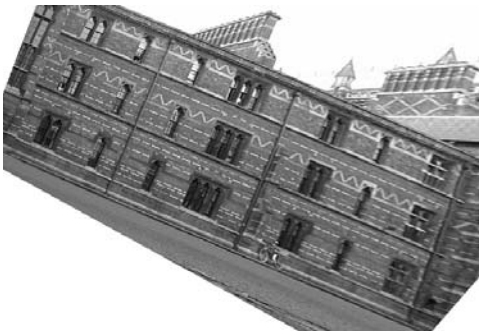


Figure 3: *A metric rectified image of the plane using only the non-metric part,* $\mathtt{N}$*, of the rectification homography. Angles and length ratios are the same as in the previous rectified image.*

orientation, origin and scale, determined by the metric component of the transformation.

Since in this case $\mathtt{H}$ is known it is, of course, unnecessary to compute the images of the circular points and thence $\mathtt{N}$. In the sequel, however, we will be presenting methods of determining the image of the circular points where four points *cannot be measured*.

The goal from hereon is to recover the image of the circular points, or equivalently the non-metric component of the rectification homography. In the following section it is shown that it is often convenient to recover these parameters in two stages.

2.2. Stratified metric rectification

In this section it is shown that rectification can be achieved without knowledge of four point correspondences, providing a technique for those cases where coordinate measurements on a world plane cannot be obtained. Geometric relationships on the world plane, such as parallelism and orthogonality, are employed to compute the projective and affine components of the homography.

The non-metric component $\mathtt{N}$ of the rectification homography can be further decomposed into two matrices:

$$\mathtt{N} = \mathtt{A}\mathtt{P}$$

The first of these is an affine transformation which encodes two of the four rectification parameters:

$$\mathtt{A} = \begin{pmatrix} \frac{1}{\beta} & -\frac{\alpha}{\beta} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{5}$$

The second component of $\mathtt{N}$ is a projective transformation

$$\mathtt{P} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ l_1 & l_2 & 1 \end{pmatrix} \tag{6}$$

The vector $\mathbf{l}_\infty = (l_1, l_2, 1)^\top$ is the vanishing line of the world plane. This is the image of the line at infinity of the world plane. It is homogeneous and has two degrees of freedom which encode all the pure projective distortion of the plane.

Under a stratified rectification scheme the two projective components, encoded by the vanishing line of the plane, are recovered first. The two affine components, corresponding to the co-ordinates of the imaged circular points on the vanishing line, are then computed. The term 'stratified' originates in the computer vision literature with the work of Koenderink[17] and Faugeras[10].

From Projective to Affine
The first stage is to determine $\mathtt{P}$, which requires identifying the vanishing line $\mathbf{l}_\infty$ of the plane. Parallel lines on the world plane intersect at vanishing points in the image, and the vanishing points lie on $\mathbf{l}_\infty$. Two or more such points determine $\mathbf{l}_\infty$, as shown in figure 4. Once $\mathtt{P}$ is determined the image can be affine rectified, as in figure 5. Note, the lines used to compute the vanishing points must be parallel to the world plane, but need not lie on that plane. Many other constraints may be used to determine the vanishing line. For example, a single set of equally spaced parallel lines on the plane is sufficient to determine $\mathbf{l}_\infty$[20]; and vanishing points can be determined from a known length ratio on a line.

Affine rectification is necessary if textures are to be acquired from images and used in imaged based modelling. Formats such as VRML specify the map between a plane and texture image by three points (an affine map). If the texture is projectively distorted then the resulting rendering of the texture mapped planes will be incorrect.

From Affine to Metric
Having recovered the plane geometry up to an affine transformation by applying the matrix $\mathtt{P}$, the final stage is the recovery of metric geometry. This requires an affine transformation of the plane, $\mathtt{A}$, that will restore angles and length ratios for non-parallel lines.

We are most concerned in this paper with rectification of building facades - where rectangular structures exist, such as the facade outline or windows (as in figure 1). Such structures generally uniquely determine the projective rectification parameters, but only provide one constraint on the two affine parameters α and β. Specifically, rectangles or two orthogonal directions such as vertical and horizontal (on the

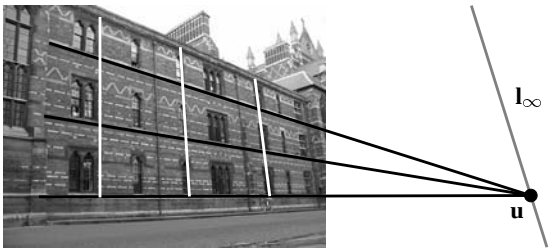


Figure 4: *The images of parallel world lines determine vanishing points: The black lines are parallel in the world and intersect in the vanishing point* $\mathbf{u}$*. The white parallel lines intersect in a vanishing point not shown because it is far from the image boundaries. The two vanishing points determine the vanishing line* $\mathbf{l}_\infty$*, and the projective rectification parameters.*

Figure 5: *The affine rectified image. Notice that parallelism is restored, but angles and ratios of lengths are still incorrect.*

world plane) provide a pair of vanishing points in orthogonal directions. This single orthogonality constraint determines the metric rectification up to a one parameter family - the ambiguity corresponding to the relative scale of vertical and horizontal directions. Figure 6 shows two of the possible rectifications. The first is computed with an arbitrary selection of relative scale of a window on the facade. This ambiguity is removed if the aspect ratio (width to height) of the window is known. Alternatively, in section 4.6 it is shown that the one parameter ambiguity can also be resolved from partial knowledge of the internal camera parameters. The stratified rectification method is summarized in Algorithm 1.

(a) (b)

Figure 6: *The aspect ratio ambiguity in relative scale of vertical and horizontal directions. (a) An incorrectly scaled image. (b) The correctly scaled image, from the known length ratio of the sides of a window. A method to resolve the ambiguity without scene measurement is given in section 4.6*

More generally, single constraints on the affine parameters may be obtained from: a known angle between lines; equality of two (unknown) angles; and, a known length ratio (details of these appear elsewhere[18]). The affine parameters are also both determined from the image of a circle on the world plane – a circle intersects the line at infinity in the two circular points (hence their names); a circle is generally imaged as an ellipse, and the ellipse intersects the vanishing line in two points, the images of the circular points.

Algorithm 1: Computing metric plane rectification.
1. Intersect parallel line segments in two orthogonal directions to obtain the vanishing points $\mathbf{u}$, $\mathbf{v}$.
2. Compute the vanishing line for the plane:

$$\mathbf{l}_\infty = (l_1,\ l_2,\ 1)^\top = \mathbf{u} \times \mathbf{v}$$

3. Rectify the vanishing points by the homography $\mathtt{P}$ to give affine plane geometry as in (6)

$$\mathbf{u}_A = \mathtt{P}\mathbf{u}\ ,\ \ \mathbf{v}_A = \mathtt{P}\mathbf{v}$$

$\mathbf{u}_A$ and $\mathbf{v}_A$ are points at infinity, and represent direction.
4. Rotate by $\mathtt{R}$ so that $\mathbf{u}_A$ is aligned with the horizontal axis. The angle between the directions of $\mathbf{u}_A$ and $\mathbf{v}_A$ is θ.
5. The affine transformation

$$\mathtt{A}_1 = \begin{pmatrix} 1 & -\cot(\theta) & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

restores metric geometry up to the unknown aspect ratio.
6. The aspect ratio is corrected by a second affine transformation

$$\mathtt{A}_2 = \begin{pmatrix} \mu & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

where μ is the correction required. The correction can be found from a ratio of lengths in the directions of $\mathbf{u}_A$ and $\mathbf{v}_A$ or using the method of section 4.6.
7. Rectify the image with the composed transformation $\mathtt{A}_2\mathtt{A}_1\mathtt{R}\mathtt{P}$

3. Single view reconstruction I

We show in this section how, given affine calibration of a reference plane, metric measurements orthogonal to the reference plane can be computed. This allows reconstruction of a ground plane and vertical walls from a single view. No explicit calibration of the camera is necessary.

3.1. Measuring distances of points from planes using one view

We describe the following result: Given the vanishing line of a reference plane, the vanishing point for directions orthogonal to the plane and a reference distance orthogonal to the plane, the orthogonal distance of any point from the plane can be computed from the image of the point and the image of the vertical intersection with the ground plane [5, 6, 19].

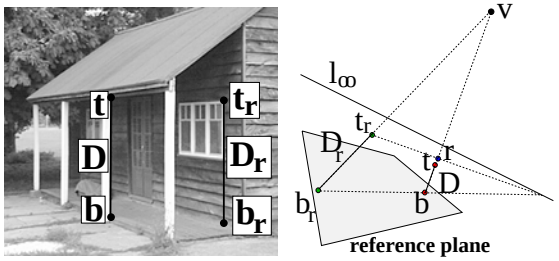
Note that no knowledge of the camera is necessary to apply the above technique. In fact, the position of the camera relative to the reference plane can also be computed.

(a) (b)

Figure 7: *An example of measuring heights from 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.*

Consider for example the image in figure 7(a). Sufficient parallel lines are present to determine the vanishing line of the ground plane and the vertical vanishing point (see figure 14(c)). The height of any object in the scene can then be computed relative to a reference using the geometry of figure 8. In this case the measured height of the top of the window is used as reference and the height of the pillars is computed. The heights are correctly found to all be the same, as shown in figure 7(b). The method is summarized in Algorithm 2.

Furthermore, if the reference plane can be metric rectified, then the 3D position of points in space and therefore a complete 3D model can be computed.



(a) (b)

Figure 8: *Notation for Algorithm 2. (a) Real image. (b) Schematic. Base points* $\mathbf{b}$ *and* $\mathbf{b}_r$ *for vertical distances lie on the ground (reference) plane. Top point* $\mathbf{t}_r$ *is at a known (or reference) distance* D_r *from the reference plane. The four aligned points* $\mathbf{b}$, $\mathbf{t}$, $\mathbf{r}$, $\mathbf{v}$ *and their corresponding world points define a line-to-line homography. Distance* D *from* $\mathbf{t}$ *to the ground is computed using Algorithm 2.*

An example of this technique is presented in figure 9, which shows an image of the painting "La Flagellazione di Cristo" by Piero della Francesca (1416 - 1492). Figures 9(b-d) show rendered views of the 3D model reconstructed from the scene.

The reconstruction is possible as a result of strict adherence to Renaissance perspective rules by the artist. The painting represents scene geometry almost exactly as it would be captured by a perspective camera, so the techniques we have described are valid. The ground plane is chosen as reference and is metric rectified from the square floor patterns as described in section 2. The vertical vanishing point follows from the intersection of the vertical lines and consequently the heights of people and columns can be computed relative to a chosen height in the scene. Once the position on the ground of each vertical object is estimated the 3D model is complete.

Figure 9(b) shows a view of the reconstructed model. Note that the people are represented simply as flat silhouettes since it is not possible to recover volume from one image. The columns have been approximated with cylinders.

A significant area of the floor is occluded by the foreground figures. However, the pattern on the floor exhibits considerable symmetry which is apparent in the rectified view, and therefore it is possible to touch up the texture map, as we have done in figures 9(c) and 9(d).

Algorithm 2: Computing distances of points from a plane:
1. Compute the reference plane vanishing line $\mathbf{l}_\infty$.
2. Compute the orthogonal vanishing point $\mathbf{v}$.
3. Notation from here on refers to figure 8. Select the top and base reference points. The top reference point $\mathbf{t}_r$ is the image of a point $\mathbf{T}_r$ off the reference plane; the base reference point $\mathbf{b}_r$ is image of the point $\mathbf{B}_r$, the orthogonal projection of $\mathbf{T}_r$ onto the reference plane.
4. Measure the world reference distance D_r between the reference points $\mathbf{T}_r$ and $\mathbf{B}_r$.
5. Select the point $\mathbf{t}$, image of the point $\mathbf{T}$ whose distance D from the plane has to be measured.
6. Select the corresponding base point $\mathbf{b}$.
7. Compute the point $\mathbf{r}$

$$\mathbf{r} = (\mathbf{v} \times \mathbf{b}) \times (\mathbf{t}_r \times (\mathbf{l}_\infty \times (\mathbf{b}_r \times \mathbf{b})))$$

8. Compute the 2×2 line-to-line homography $\mathtt{Q}$ as

$$\mathtt{Q} = \begin{pmatrix} D_r\,(d(\mathbf{v},\mathbf{b}) - d(\mathbf{r},\mathbf{b})) & 0 \\ -d(\mathbf{r},\mathbf{b}) & d(\mathbf{v},\mathbf{b})d(\mathbf{r},\mathbf{b}) \end{pmatrix}$$

where $d(\mathbf{p}_1, \mathbf{p}_2)$ is distance between two image points $\mathbf{p}_1$ and $\mathbf{p}_2$.
9. Compute the world distance D of the input point $\mathbf{T}$ from the reference plane as

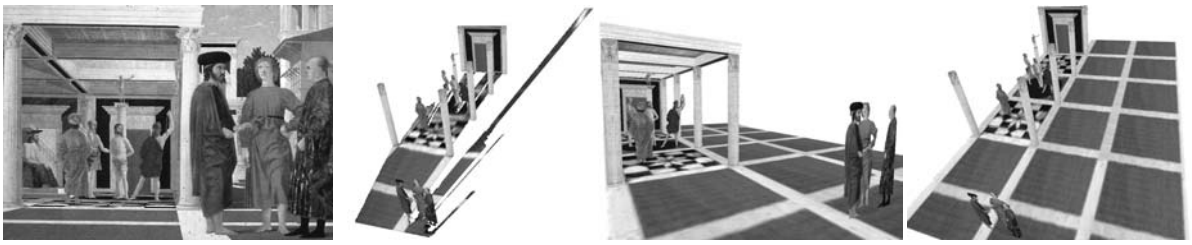
$$D = s_1 / s_2$$

where s_1 and s_2 are components of the 2-vector $\mathbf{s}$ defined by

$$\mathbf{s} = \mathtt{Q} \begin{pmatrix} d(\mathbf{t},\mathbf{b}) \\ 1 \end{pmatrix}$$

4. Camera calibration

There are several reasons why camera calibration, i.e. determining the internal parameters (interior orientation) of a camera, is an important step in single view reconstruction. The first is that it allows complete rectification of planes



(a) (b) (c) (d)

Figure 9: *La Flagellazione di Cristo. (a) The painting (1460, Urbino, Galleria Nazionale delle Marche). (b) A view of the reconstructed 3D model where the roof has been removed and the floor is partly occluded by the foreground figures. People are represented as flat silhouettes and the columns have been approximated with cylinders. (c) A view of the 3D model where the floor has been touched up by making use of the symmetry of its pattern. The partially seen ceiling has been reconstructed too. (d) Another view of the model with the roof removed to show the relative positions of people and columns in the scene. Notice the repeated geometric pattern on the floor in the area delimited by the columns. These figures appear in the colour section as figures 20 and 21.*

given only their vanishing line. Secondly, with known internal parameters and vanishing line, the orientation of a plane relative to the camera may be computed; and thirdly from two vanishing lines the relative orientation between two planes can be computed.

Furthermore we will show that a camera may be calibrated using the type of rectification constraints (parallel lines, orthogonality) described in the previous section for planes. Again, it is not necessary for scene measurements to be available. Thus in contrast to standard photogrammetry methods for camera calibration[22, 25], where 3D points in a world co-ordinate system are required, we will employ the constraints which are plentifully available in architectural scenes. Indeed, it will be shown that a camera can be calibrated directly from the rectification parameters of a set of planes. First we define the camera model and notation.

4.1. Camera model

The camera calibration matrix is specified by a five parameter upper triangular matrix

$$\mathtt{K} = \begin{pmatrix} f & k & u_0 \\ 0 & rf & v_0 \\ 0 & 0 & 1 \end{pmatrix} \tag{7}$$

An image point $\mathbf{x}$ is related to a point in the camera's coordinate system $\mathbf{x}_c$ as $\mathbf{x} = \mathtt{K}\mathbf{x}_c$.

Parameter f is the focal length of the camera. The aspect ratio of the camera r depends on the relative scaling of the vertical and horizontal camera axes. The line from the camera centre perpendicular to the image intersects the image at the principal point with co-ordinates $(u_0, v_0)^\top$. The skew, k, is a factor dependent on the physical angle θ between the u and v axes in the sensor array, given by $k = f\cot(\theta)$. Note that radial lens distortion is ignored, but can be corrected in cases where it is significant[8].

In many cases a simplified camera model may be used. A CCD camera, for example, has zero skew ($k = 0$) and unit aspect ratio ($r = 1$). The resultant simplified or *natural camera* is

$$\mathtt{K} = \begin{pmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{pmatrix} \tag{8}$$

The more general camera model (7) does apply in certain situations however. For example if a photographic negative is enlarged, and the paper is not parallel to the plane of the negative. In many situations the principal point is located near the centre of the image, and often can be approximated by the image centre. However, it cannot be assumed that this is always the case because photographs (and images) are sometimes cropped before display.

Before describing the method of computing internal parameters from orthogonal vanishing points and rectified planes, we require one more shot of projective geometry, which is an understanding of the image of the absolute conic. The applications are explored in the sections that follow.

4.2. The image of the absolute conic

The absolute conic is an esoteric entity lying on the plane at infinity in 3D. Its image is important here because it provides a simple object for reasoning about orthogonality in the image and is simply related to the camera calibration.

The image of the absolute conic is a conic ω defined as[10]

$$\omega = \mathtt{K}^{-\top}\mathtt{K}^{-1} \tag{9}$$

It is a homogeneous 3×3 symmetric matrix with 5 degrees of freedom – the 6 independent elements of the matrix less one for overall scale. A point $\mathbf{x}$ lying on the conic satisfies[21]

$$\mathbf{x}^\top \omega \mathbf{x} = 0$$

Points on ω are complex, but this presents no difficulty since the matrix itself is real valued. A conic is defined by five points.

The calibration matrix $\mathtt{K}$ may be computed from ω by Cholesky decomposition[13], a factorisation method that decomposes a symmetric matrix into the product of a lower triangular and upper triangular matrix of the form of (9). Determining ω in an image, then, is equivalent to knowing the camera internal parameters.

There are two important properties of ω: first, the imaged circular points of any plane lie on ω — thus each rectified plane provides two points on ω and so provides two of the five constraints necessary to determine ω; second, ω determines orthogonality of rays back projected from image points. A pair of vanishing points $\mathbf{u}$ and $\mathbf{v}$ arising from orthogonal directions in the world satisfy

$$\mathbf{u}^\top \omega \mathbf{v} = 0 \tag{10}$$

and are said to be conjugate with respect to ω[27].

We now have a geometric entity with which to reason about the internal parameters of the camera and their relationship to constraints between scene objects such as planes. It will be seen that these relationships are represented concisely and are linear.

4.3. Calibration from orthogonal vanishing points

The image of the absolute conic is a matrix with elements

$$\omega = \begin{pmatrix} \omega_1 & \omega_2 & \omega_4 \\ \omega_2 & \omega_3 & \omega_5 \\ \omega_4 & \omega_5 & \omega_6 \end{pmatrix} \tag{11}$$

Writing $\mathbf{u} = (u_1,\ u_2,\ u_3)^\top$ and $\mathbf{v} = (v_1,\ v_2,\ v_3)^\top$, (10) takes the form

$$\begin{aligned} &u_1v_1\omega_1 + (u_1v_2 + u_2v_1)\omega_2 + u_2v_2\omega_3 \\ &+(u_1v_3 + u_3v_1)\omega_4 + (u_2v_3 + u_3v_2)\omega_5 + u_3v_3\omega_6 = 0 \end{aligned} \tag{12}$$

Writing the elements of ω as a vector

$$\omega_v = (\omega_1,\ \omega_2,\ \omega_3,\ \omega_4,\ \omega_5,\ \omega_6)^\top$$

and the coefficients of the elements of ω_v in (12) as $\kappa_{uv} = (u_1v_1,\ u_1v_2 + u_2v_1,\ u_2v_2,\ u_1v_3 + u_3v_1,\ u_2v_3 + u_3v_2,\ u_3v_3)^\top$; then (12) becomes

$$\kappa_{uv}{}^\top \omega_v = 0 \tag{13}$$

This is a *linear* constraint on the 5 parameters of ω. For each pair of orthogonal vanishing points, an additional constraint of this form is obtained. Five such constraints determines ω and thence $\mathtt{K}$. The five constraints are a simple linear system that may be written as a 5×6 matrix, and ω_v computed as the null-vector of this matrix. $\mathtt{K}$ follows from ω by Cholesky decomposition.

A typical real world application such as an image of a building provides three orthogonal directions, and thus three constraints on ω. Therefore there are insufficient constraints to solve for the full five parameter model, but there are sufficient to determine the three parameter natural camera model (8) if this is applicable. The extra camera constraints of zero skew and unit aspect ratio provide additional constraints on ω.

The constraints provided by the fact that the camera is natural are that the known circular points of the image plane lie on ω. The image plane is thus treated as a rectified plane (see section 4.4). Equivalently, if we expand the elements of ω in terms of the parameters of $\mathtt{K}$, it is apparent that zero skew and unit aspect ratio imply

$$\omega_2 = 0 \quad \text{and} \quad \omega_1 - \omega_3 = 0 \tag{14}$$

The camera computation algorithm applicable to three orthogonal vanishing points is summarized in Algorithm 3. As shown by Caprile and Torre[3], the orthogonality equations for a natural camera are equivalent to a simple construction: the triangle with the three orthogonal vanishing points as vertices has the principal point as its orthocentre. Figure 10 (a) shows an image of a building with lines in three orthogonal directions. The vanishing points of each of these three directions, shown in figure 10 (b), provide the three constraints on the internal parameters, and define the triangle with principal point at its orthocentre.

Algorithm 3: Computing the internal parameters from three orthogonal vanishing points:

1. Intersect parallel line segments in three orthogonal directions to obtain the vanishing points $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$.
2. Compute the coefficient matrix $\mathtt{A}$ from (13) and (14):

$$\mathtt{A}^\top = \begin{pmatrix} u_1v_1 & u_1w_1 & v_1w_1 & 0 & 1 \\ u_1v_2 + u_2v_1 & u_1w_2 + u_2w_1 & v_1w_2 + v_2w_1 & 1 & 0 \\ u_2v_2 & u_2w_2 & v_2w_2 & 0 & -1 \\ u_1v_3 + u_3v_1 & u_1w_3 + u_3w_1 & v_1w_3 + v_3w_1 & 0 & 0 \\ u_2v_3 + u_3v_2 & u_2w_3 + u_3w_2 & v_2w_3 + v_3w_2 & 0 & 0 \\ u_3v_3 & u_3w_3 & v_3w_3 & 0 & 0 \end{pmatrix}$$

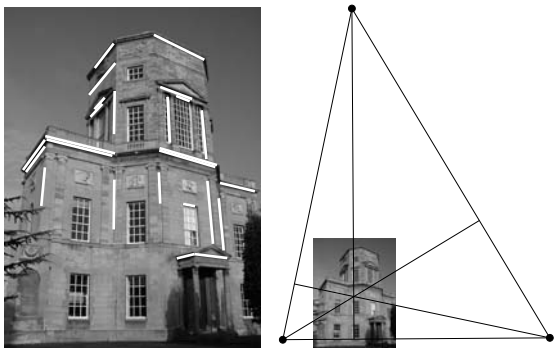
3. Compute ω_v as a null vector:

$$\mathtt{A}\omega_v = \mathbf{0}$$

4. Form the symmetric matrix ω from ω_v as in (11).
5. Compute the Cholesky decomposition: $\omega = \mathtt{G}\mathtt{G}^\top$
6. $\mathtt{K} = \mathtt{G}^{-\top}$

Note: if ω is negative definite, it must be rescaled by -1 before Cholesky decomposition. If it is neither positive definite or negative definite, it will not decompose to a real matrix, and indicates an error.

Note that in cases where parallel line segments are imaged parallel, that is the vanishing point is at infinity, degeneracies occur. The nature of these degeneracies is regrettably beyond the scope of this paper, and a warning will have to suffice.



(a) (b)

Figure 10: *Internal parameter estimation. (a) Image of the Radcliffe Observatory, Oxford, with sets of parallel line segments defining vanishing points of three orthogonal directions. (b) The triangle with the vanishing points as vertices. The principal point of the camera lies at the orthocentre of the triangle. f = 1041.3, u_0 = 384.1 and v_0 = 543.3. The image size is 768 by 1024.*

4.4. Rectified planes

If the four rectification parameters for a scene plane are computed, then the imaged circular points for that plane are known. These points lie on ω and, for $\mathbf{I} = (\alpha - i\beta, 1, -l_2 - \alpha l_1 + i l_1\beta)^\top$

$$\mathbf{I}^\top \omega \mathbf{I} = 0$$

The real and imaginary parts of $\mathbf{I}$ give

$$\begin{aligned}(\beta^2-\alpha^2)\omega_1 - 2\alpha\omega_2 - \omega_3 + 2(l_1(\alpha^2-\beta^2)+\alpha l_2)\omega_4 & \\ +2(\alpha l_1 + l_2)\omega_5 + (l_1^2\beta^2 - (\alpha l_1 + l_2)^2)\omega_6 &= 0 \\ 2\alpha\beta\omega_1 + 2\beta\omega_2 - 2(\beta l_2 + 2\alpha\beta l_1)\omega_4 - 2\beta l_1\omega_5 & \\ +2(\alpha\beta l_1^2 + \beta l_1 l_2)\omega_6 &= 0\end{aligned}$$

Each rectified plane thus provides two linear constraints on ω. We may, of course, write the constraints in the matrix form used for orthogonal vanishing points above, and combine constraints from both sources.

With three rectified planes there are six points on ω and the camera internal parameters are over constrained. With only two rectified planes, four points on ω are known, and thus there are four constraints on ω. Additional constraints on aspect ratio *or* skew may then be included exactly as in the previous section. Note that the world planes need not be orthogonal.

4.5. Rectification with known internal parameters

Given the internal parameters of the camera $\mathtt{K}$, any plane for which the vanishing line is known can be rectified[16]. This is because the vanishing line of a given plane intersects ω in the circular points corresponding to that plane. So, the circular points can be computed from the intersection of ω and $\mathbf{l}_\infty$ as shown below, and thence the metric rectification is determined as in section 2. The rectification ambiguity described in section 2.2 can thus be resolved if the camera is known. Note, that the rectification is a simple linear procedure and does not require computation of rotation matrices.

The circular points at the intersection of ω and $\mathbf{l}_\infty$ may be calculated from the first component of $\mathbf{I} = (I_1,\ I_2,\ I_3)^\top = (\alpha - i\beta, 1, -l_2 - \alpha l_1 + i l_1\beta)^\top$ by solving the quadratic

$$\begin{aligned}(1 + 2u_0 l_1 + l_1^2(u_0^2+v_0^2+f^2))I_1^2 & \\ +2(l_2 u_0 + l_1 v_0 + l_1 l_2(u_0^2+v_0^2+f^2))I_1 & \\ +2l_2 v_0 + l_2^2(u_0^2+v_0^2+f^2) + 1 &= 0\end{aligned} \tag{15}$$

for the case of a natural camera. The affine parameters are then the real and imaginary parts of I_1.

The vanishing line and camera also determine the orientation of the world plane relative to the camera[4]. The normal to the plane in camera centred co-ordinates $\mathbf{x}_c$ is

$$\mathbf{n} = \mathtt{K}^\top \mathbf{l}_\infty$$

where $\mathbf{l}_\infty$ is the vanishing line of the plane.

The relative orientations of two planes may be computed from their vanishing lines as

$$\cos(\theta) = \frac{{\mathbf{l}_{\infty 1}}^\top \omega^{-1} \mathbf{l}_{\infty 2}}{({\mathbf{l}_{\infty 1}}^\top \omega^{-1} \mathbf{l}_{\infty 1})^{1/2} ({\mathbf{l}_{\infty 2}}^\top \omega^{-1} \mathbf{l}_{\infty 2})^{1/2}} \tag{16}$$

with θ the angle between the planes.

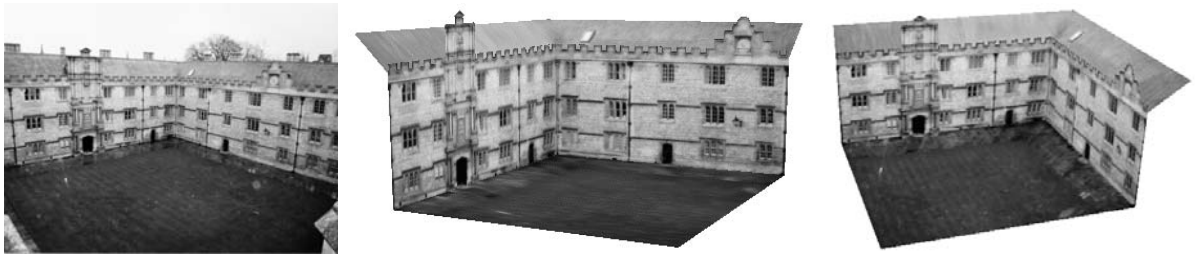
4.6. Single plane rectification with partial internal parameters

It is common to require the rectification of a plane where the aspect ratio ambiguity of section 2.2 exists, but there are no other constraints available from the scene from which to compute the camera calibration using a technique such as the vanishing point method above. However, the ambiguity can be resolved for a natural camera for which the principal point is known approximately, for example as the image centre. The only remaining internal parameter is then f.

Consider figure 11. The building facade has two dominant directions which are orthogonal. These provide three of the four plane rectification parameters and result in a rectification with an aspect ratio ambiguity. The internal parameters of the camera are also not fully constrained, since only one orthogonal pair of vanishing points is present to apply (10). However, using a natural camera with principal point at the centre of the image, f can be computed from the single constraint, and hence $\mathtt{K}$ is also fully determined. A complete rectification for the planes can be computed, and is shown in figure 11 (b).

5. Single view reconstruction II

The goal of this section is to build 3D models which require the rectification of 2 or more planes. In general, the task is to



(a) (b) (c)

Figure 12: *3D reconstruction from a single image. (a) Fellows quad, Merton College, Oxford. (b) and (c) Views of the 3D model created from the single image. The vanishing line of the roof planes is computed from the repetition of the texture pattern[20]. The figures appear in the colour section as figure 22.*

(a) (b)

Figure 11: *Plane rectification via partial internal parameters. (a) Original image. (b) Rectification where the relative scaling of vertical and horizontal directions assumes a natural camera with the principal point at the centre of the image. The focal length is computed from the single orthogonal vanishing point pair. Measurement of the aspect ratio of a window indicates a difference of 3.7% between true and computed values. Note that the two parallel planes, the upper building facade and the lower shopfront, are both correctly rectified, but scene planes not parallel to these two are distorted by the rectification homography. This distortion is visible on the area of overhang of the upper facade.*

reconstruct a scene from recognisable scene primitives such as lines, planes and spheres by computing their spatial layout. The plane rectification techniques we have presented are ideal for the task of reconstructing models of buildings that have planar surfaces, so we will restrict the discussion to planes.

In practice it is often not possible to completely rectify any planes until the camera calibration has been estimated, using, for example, the vanishing point method of section 4.3. One common situation is where there are two or three mutually orthogonal planes with lines in orthogonal directions on the planes, as in figure 12(a). There are three dominant planes in the scene; the building facades on the left and right and the ground plane. The parallel line sets in three orthogonal directions define three vanishing points and thus the natural camera may be computed. From the vanishing lines of the three planes, likewise determined by the vanishing points, and ω, we can rectify each of the planes (from (15)). Another situation that often arises is where the rectification of two planes can be computed from the scene. This then determines the natural camera (as in section 4.4).

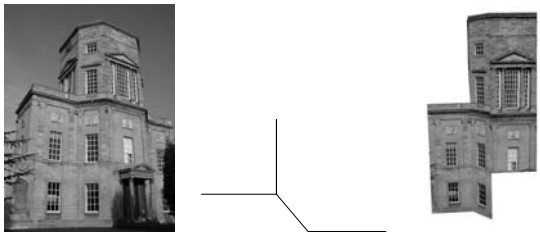
Having computed the camera, the relative *orientation* of planes in the scene that are not orthogonal can be computed if their vanishing lines can be found. Their relative *positions* and *dimensions* can be determined if the intersection of a pair of planes is visible in the image, so that there are points common to both planes. Relative size can be computed from the rectification of a distance between common points using the homographies of both planes.

There are two ways of proceeding: one is to sequentially build the model from a reference plane, much like building a staircase from the bottom up. There is of course a problem with accumulated error in this sequential approach; a second approach is where all the planes intersect the reference. Then accumulated error can be avoided because all the planes can be specified relative to the rectified reference.

Taking the left facade as reference in figure 12(a), its correctly proportioned width and height are determined by the rectification. The right facade and ground planes define 3D planes orthogonal to the reference (we have assumed the orthogonality of the planes in computing the camera, so relative orientations are defined). Scaling of the right and ground planes is computed from the points common to the planes and this completes a three orthogonal plane model. Reconstruction of the roof planes is achieved by determining their orientation from their respective vanishing line. Views of the model, with texture mapped correctly to the planes, appear in figures 12(b) and (c). The method is summarized in Algorithm 4.

A second example appears in figure 13. Four planes are visible in the scene, three of which are orthogonal. In addition, sufficient parallel line sets are available to determine camera internal parameters and vanishing lines for each plane. The planes can thus all be rectified, and the relative orientation of the non-orthogonal plane determined. The re-

construction appears in plan view in figure 13(b), and a view of the model in 13(c).



(a) (b) (c)

Figure 13: *Single view reconstruction. (a) Original image. (b) Plan view of the 3D model, showing the four planes reconstructed. (c) A view of the textured reconstruction. The angle between the non-orthogonal plane and the facade planes has been computed from (16) to be 46^o. The true angle is 45^o.*

Algorithm 4: Computing a 3D reconstruction of a single view of two or three mutually orthogonal planes:
1. Intersect parallel line segments in three orthogonal directions to obtain the vanishing points **u**, **v** and **w**.
2. Compute K as in Algorithm 3.
3. Compute vanishing lines for each plane from two vanishing points, eg

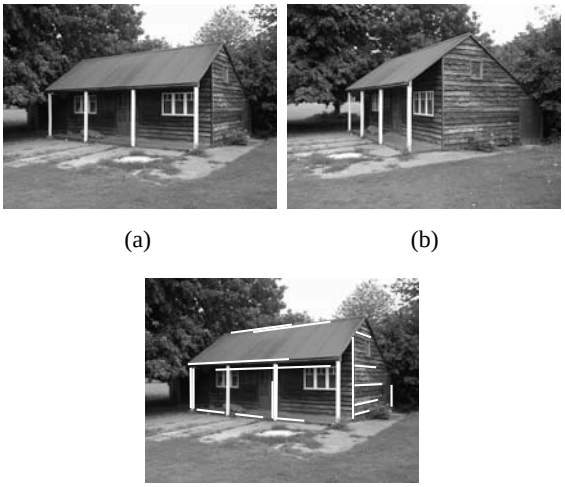
$$\mathbf{l}_\infty = (l_1,\ l_2,\ 1)^\top = \mathbf{u} \times \mathbf{v}$$

4. Compute affine rectification parameters from (15) and planar rectification homographies as in (3).
5. Select a reference plane and construct the corresponding model plane such that it is has rectified dimensions.
6. Compute relative orientations of rectified planes from (16). From the common points with the other planes and the rectification homographies of the planes, compute relative scale factors.

6. Reconstruction from two views

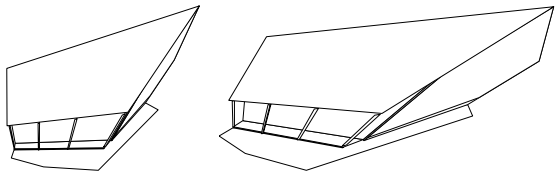
If the internal parameters and pose of two or more cameras is unknown, a reconstruction from matched scene features is possible up to a homography of 3D[9, 14]. This means that the initial reconstruction exhibits 3D projective deformation, so that parallel lines do not appear parallel, angles are incorrect, and so on. From the two views in figure 14, for example, a projective reconstruction can be computed, and is shown as a wireframe model in figure 15.

The homography relating the projectively distorted model of the scene to a metric reconstruction can be decomposed in an identical manner to the planar projective distortion case described in section 2. If the similarity component is disregarded an eight parameter rectification homography remains, three projective parameters corresponding to the plane at infinity and five affine parameters describing the absolute conic[11]. Again, we wish to employ parallel and orthogonality relationships in the scene to compute first projective and then affine rectification parameters. Space limitations mean that this method can only be sketched.



(a) (b) (c)

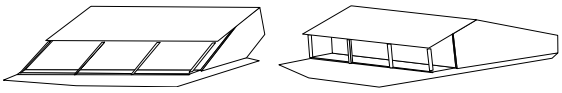
Figure 14: *Images used in a two view reconstruction. (a) and (b) two views of the scene. (c) Some of the parallel line segments in the first view. These define three vanishing points for which there are corresponding vanishing points in the second view.*



(a) (b)

Figure 15: *Two views of the projective reconstruction computed from the image pair of figure 14.*

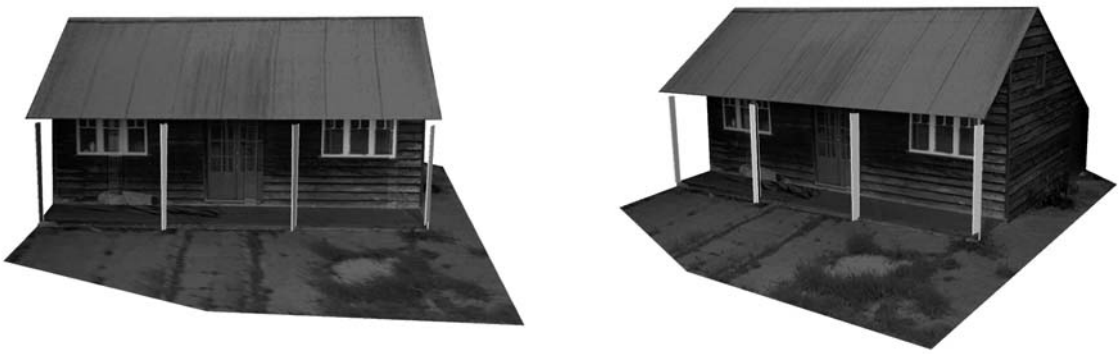
A 3D point on the plane at infinity is computed by triangulation for corresponding vanishing points in the two images. Three such points determines the plane. Once the plane at infinity is computed, the structure can be affine rectified, as shown in figure 16, where parallelism is restored.



(a) (b)

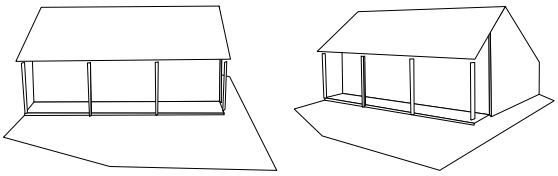
Figure 16: *Two views (orthographic projections) of the affine rectified structure. The affine rectification is determined from three sets of parallel lines in the scene.*

The five affine rectification parameters associated with the



(a) (b)

Figure 18: *Texture mapped views of metric rectified structure.*



(a) (b)

Figure 17: *Two views of the metric rectified structure. The metric rectification is determined from the orthogonality of the three vanishing point directions, together with the known aspect ratio and zero skew of the camera.*

absolute conic must now be found. If the vanishing points are orthogonal, we can apply orthogonality constraints on the absolute conic in 3D using the directions defined by the vanishing points. These constraints are identical in form to those applied to the image of the absolute conic ω in section 4.3. With three orthogonal vanishing points and the natural camera constraints of zero skew and unit aspect ratio, the affine rectified structure can be metric rectified. As shown in figures 17 and 18, parallelism, angles and relative lengths are all correct. Texture is mapped onto the model planes from the most appropriate image.

7. Implementation details

Line segments are detected by: Canny edge detection at sub-pixel accuracy[2]; edge linking; segmentation of the edgel chain at high curvature points; and finally, straight line fitting by orthogonal regression to the resulting chain segments.

Due to 'noise' a set of imaged parallel line segments will generally *not* intersect in a point. Often the vanishing point is then computed by finding the closest point to all the measured lines. However, this is not optimal. The maximum likelihood estimate (MLE) of the vanishing point is found by fitting a set of lines that do intersect in a single point, and which minimise the sum of squared orthogonal errors from the endpoints of the measured line segments, as shown in figure 19. The MLE is computed by non-linear minimisation.

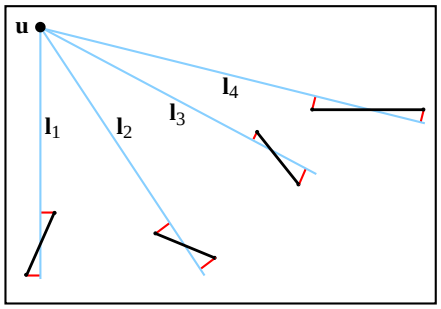


Figure 19: *MLE vanishing point estimation: The vanishing point* $\mathbf{u}$ *is estimated as the intersection of the fitted lines* $\mathbf{l}_i$ *(in gray), which minimise the orthogonal distances from the endpoints of measured imaged parallel line segments (shown in black).*

8. Conclusions

We have demonstrated methods which correctly model and take account of perspective distortion – though of course the methods work equally well if there is little or no perspective distortion of the scene. There are a number of extensions to this work:

1. If two or more views of an object are available, but under conditions which prevent computation of corresponding points matches, single view reconstructions can be created and joined. This might be necessary, for example, when two views of an object have very limited overlap.
2. Although we have concentrated on planes, other surfaces

may be reconstructed from a single view in a similar manner, for example parametrized surfaces, such as quadrics (e.g. spheres for domes, cylinders for columns).
3. Surfaces which repeat (e.g. by symmetry) in a single view may be fully reconstructed. This is because the single view of the surface is equivalent to two images of the symmetric half of the surface, each image from a different viewpoint. This means that two view reconstruction methods can be employed in a single view.
4. Once the camera is calibrated using orthogonality of vanishing points and rectification of planes, it may then be used in the reconstruction of other surfaces.
5. For an optimal reconstruction the final 3D model should be 'polished' by minimizing reprojection errors in the image subject to the parallelism and orthogonality constraints obtained from the scene.

Acknowledgements

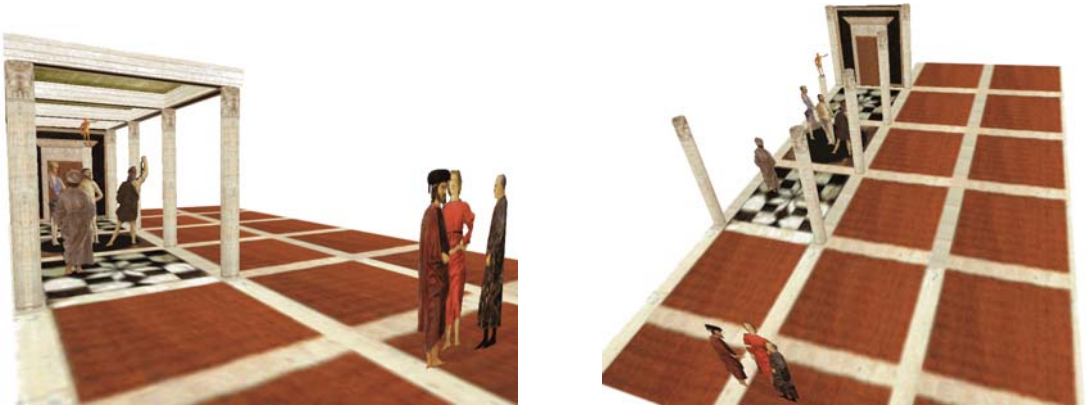
The authors would like to thank Dr Andrew Fitzgibbon, Dr Ian Reid and Frederick Schaffalitzky for their assistance in the preparation of this paper. We are grateful to EU Esprit project Improofs for financial support.

References

1. P. Beardsley, P. Torr, and A. Zisserman. 3D model acquisition from extended image sequences. In *Proc. ECCV*, LNCS 1064/1065, pages 683–695. Springer-Verlag, 1996.
2. J. Canny. A computational approach to edge detection. *IEEE T-PAMI*, 8(6):679–698, 1986.
3. B. Caprile and V. Torre. Using vanishing points for camera calibration. *IJCV*, pages 127–140, 1990.
4. R. T. Collins and J. R. Beveridge. Matching perspective views of coplanar structures using projective unwarping and similarity matching. In *Proc. CVPR*, 1993.
5. A. Criminisi, I. Reid, and A. Zisserman. Computing 3d euclidean distance from a single view. Technical Report OUEL 2158/98, Dept. Engineering Science, University of Oxford, Parks Road, Oxford OX1 3PJ, U.K., 1998.
6. A. Criminisi, A. Zisserman, L. Van Gool, Bramble S., and D. Compton. A new approach to obtain height measurements from video. In *Proc. of SPIE, Boston, Massachussets, USA*, volume 3576, 1-6 November 1998.
7. P. E. Debevec, C. J. Taylor, and J. Malik. Modeling and rendering architecture from photographs: A hybrid geometry- and image- based approach. In *Proceedings, ACM SIGGRAPH*, pages 11–20, 1996.
8. F. Devernay and O. Faugeras. Automatic calibration and removal of distortion from scenes of structured environments. In *SPIE*, volume 2567, San Diego, CA, July 1995.
9. O. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In *Proc. ECCV*, LNCS 588, pages 563–578. Springer-Verlag, 1992.
10. O. D. Faugeras. Stratification of three-dimensional vision: projective, affine, and metric representation. *J. Opt. Soc. Am.*, A12:465–484, 1995.
11. O.D. Faugeras, S. Laveau, L. Robert, G. Csurka, and C. Zeller. 3-D reconstruction of urban scenes from sequences of images. Tech. report, INRIA, 1995.
12. J. D. Foley, A. Van Dam, S. K. Feiner, and J. F. Hughes. *Computer Graphics: Principles and Practice*. Addison-Wesley, 1990.
13. G. H. Golub and C.F. Van Loan. *Matrix Computations*. The John Hopkins University Press, Baltimore, MD, second edition, 1989.
14. R. I. Hartley, R. Gupta, and T. Chang. Stereo from uncalibrated cameras. In *Proc. CVPR*, 1992.
15. Y. Horry, K. Anjyo, and K. Arai. Tour into the picture: Using a spidery mesh interface to make animation from a single image. In *SIGGRAPH*, pages 225–232, 1997.
16. G.E. Karras, P. Patias, and E. Petsa. Experiences with rectification of non-metric digital images when ground control is not available. In *CIPA XV International Symposium*, 1993.
17. J. J. Koenderink and A. J. van Doorn. Affine structure from motion. *J. Opt. Soc. Am. A*, 8(2):377–385, 1991.
18. D. Liebowitz and A. Zisserman. Metric rectification for perspective images of planes. In *Proc. CVPR*, pages 482–488, June 1998.
19. M. Proesmans, T. Tuytelaars, and L.J. Van Gool. Monocular image measurements. Technical Report Improofs-M12T21/1/P, K.U.Leuven, 1998.
20. F. Schaffalitzky and A. Zisserman. Geometric grouping for automatic vanishing point and line detetion. *To appear in IVC*, 1999.
21. J. Semple and G. Kneebone. *Algebraic Projective Geometry*. Oxford University Press, 1979.
22. C. Slama. *Manual of Photogrammetry*. American Society of Photogrammetry, Falls Church, VA, USA, 4th edition, 1980.
23. R. Szeliski and S. Heung-Yeung. Creating full view panoramic image mosaics and environment maps. In *SIGGRAPH*, 1997.
24. C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization approach. *IJCV*, 9(2):137–154, November 1992.
25. R. Tsai. An efficient and accurate camera calibration technique for 3D machine vision. In *Proc. CVPR*, 1986.
26. C. Zeller. *Projective, Affine and Euclidean Calibration in Compute Vision and the Application of Three Dimensional Perception*. PhD thesis, RobotVis Group, INRIA Sophia-Antipolis, 1996.
27. A. Zisserman, D. Liebowitz, and M. Armstrong. Resolving ambiguities in auto-calibration. *Phil. Trans. R. Soc. Lond. A*, 356(1740):1193–1211, 1998.

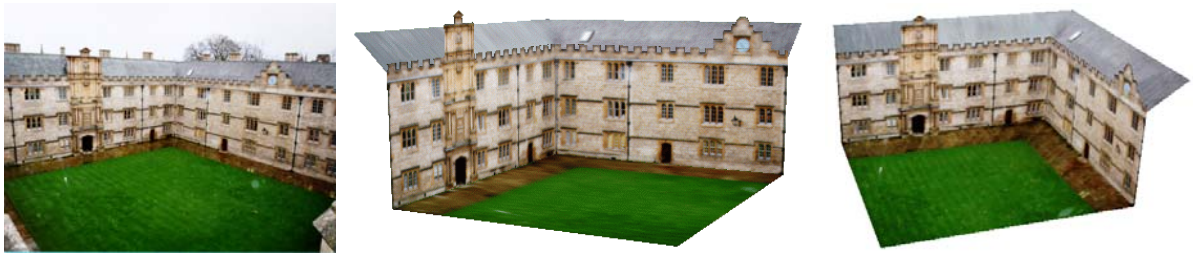
(a) (b)

Figure 20: *3D model reconstructed from a painting. (a) "La Flagellazione di Cristo" by Piero della Francesca (1460, Urbino, Galleria Nazionale delle Marche). (b) A 3D model reconstructed from the painting. The blank areas on the ground plane are occluded by the foreground figures.*



(a) (b)

Figure 21: *Further views of the model with the floor touched up in order to deal with the occlusions caused by the foreground figures. Note the texture map of the ground plane, showing the tile pattern between the pillars barely visible in the original image.*



(a) (b) (c)

Figure 22: *3D model reconstructed from a single image. (a) Fellows Quad, Merton College, Oxford. (b) and (c) Two views of the 3D model reconstructed from the single image.*