Building Architectural Models from Many Views Using Map Constraints

D.P. Robertson and R. Cipolla

University of Cambridge, Department of Engineering
Trumpington Street, Cambridge CB2 1PZ.
dpr20@eng.cam.ac.uk, cipolla@eng.cam.ac.uk

Abstract. This paper describes an interactive system for creating geometric models from many uncalibrated images of architectural scenes. In this context, we must solve the structure from motion problem given only few and noisy feature correspondences in non-sequential views. By exploiting the strong constraints obtained by modelling a map as a single affine view of the scene, we are able to compute all 3D points and camera positions simultaneously as the solution of a set of linear equations. Reconstruction is achieved without making restrictive assumptions about the scene (such as that reference points or planes are visible in all views). We have implemented a practical interactive system, which has been used to make large-scale models of a variety of architectural scenes. We present quantitative and qualitative results obtained by this system.

1 Introduction

Despite much progress [1,15,21] in the development of completely automatic techniques for obtaining geometric models from images, the best-looking architectural models are still produced interactively [6,5,17]. Interactive approaches exploit the user's higher-level knowledge to solve the difficult problems of identifying geometrically important features and wide-baseline matching. Existing interactive approaches have used one or relatively few calibrated or uncalibrated views in order to build models of a few buildings. In this paper, we address some of the problems associated with making much larger models from an arbitrarily large number of uncalibrated views of many buildings.

1.1 Previous Work

Given feature correspondences, the optimal solution for camera parameters and structure may be determined using bundle adjustment [18], which is used to distribute back projection error optimally across all feature measurements. In the context of an interactive system, we must solve the structure from motion problem given only few and noisy feature correspondences defined in images obtained from sparse viewpoints. This presents two problems: (i) bundle adjustment will only succeed provided a sufficiently good initial guess can be obtained and (ii)

A. Heyden et al. (Eds.): ECCV 2002, LNCS 2351, pp. 155–169, 2002.

even an ‘optimal’ reconstruction is no guarantee of a subjectively good-looking model.

One solution to the problem of obtaining an initial guess is to estimate camera parameters and structure simultaneously as the solution of a linear equation. Although image coordinates have a non-linear relationship with 3D coordinates under perspective projection, it is well known [6,17,16] that feature correspondences allow a linear solution for structure and camera positions if camera calibration and orientation are known. In [6] and [17] camera orientation is determined for calibrated cameras from prior knowledge of line directions. In [12] and [5] camera calibration and orientation are determined simultaneously from three vanishing points corresponding with three orthogonal directions. This approach allows a two-view initialisation of a Euclidean frame but does not address the problem of how extra views should be registered in that frame, unless we make the restrictive assumption that three vanishing points corresponding with known and orthogonal directions are visible in *all* images. Furthermore, this approach has the disadvantage that camera calibration cannot be determined for the degenerate (but very common) case of vanishing points lying at (or near) infinity in the image plane, e.g. in a photograph of a wall parallel to the image plane. [16] generalises the concept in [5,12] to any three points lying on a reference plane but relies on the almost equally restrictive assumption that four points on the reference plane are visible in all images (or at least that the reference plane is visible in all images).

Another solution is to register uncalibrated cameras sequentially [1]. A two-view initialisation defines a projective frame via the fundamental matrix. A partial reconstruction may be computed within this frame using feature correspondences. Then additional views are registered one at a time using the Discrete Linear Transformation [20]. Having determined the projection matrix for an additional view, structure may be computed for all correspondences defined in two or more views and the partial reconstruction is extended.

Whilst this approach is effective in the context of tracked features in video frames [1,8], it has severe limitations in the context of interactive systems:

1. Given only a few noisy feature correspondences partial reconstructions are likely to be quite inaccurate. Bundle adjustment may be used to improve estimated camera parameters and structure but this approach often fails in practice due to convergence to a local minima. In any case, carrying out bundle adjustment after the addition of each subsequent viewpoint is very computationally expensive.
2. Accurate camera registration by DLT depends on the accuracy of the partial reconstruction. Some reconstructed points may be quite degenerate with respect to included views and therefore inaccurate. Such points may severely compromise the accuracy of the DLT.
3. Some viewpoints may be degenerate with respect to the partial reconstruction. DLT requires at least 6 points, two of which must be non-coplanar with the remainder.

4. It is difficult to adapt the sequential approach to non-sequential image data. Given the likely inaccuracy of the partial solution and possible degeneracy of successive viewpoints with respect to that partial reconstruction, it is not clear in which order successive viewpoints should be registered within our euclidean (or projective) frame.

The second problem associated with few and noisy feature correspondences is that of obtaining a sufficiently accurate reconstruction. Architectural scenes typically contain a large number of parallel and perpendicular elements and it is subjectively very important that these relationships should be preserved as far as possible in the final model. However, small errors associated with the registration of nearby viewpoints may accumulate throughout a large set of images such that absolute errors become large. This problem may be particularly severe in cases where it is impossible to obtain images from a suitably wide range of viewpoints, e.g. a city street.

In order to address this problem, some interactive systems constrain the reconstruction process by exploiting the user's higher-level knowledge about parallelism and orthogonality. In [17] for example, scene structure is determined subject to constraints on (known) line directions and plane normals. This type of approach has the considerable disadvantage that it is not directly extensible to data sets comprising images of buildings with unknown and different orientations. Debevec et al [6] describe a system that allows the user to parameterise the scene in terms of primitives: simple geometric building blocks such as cuboids and prisms that can be combined to make more complex models. Such systems may produce excellent results but not all scenes can be expressed so simply in terms of a few geometric building blocks. Furthermore it is not always possible to find viewpoints such that a sufficiently large proportion of each primitive is visible in any one image.

1.2 Approach

We are concerned with modelling large architectural scenes. In this context, it is not always possible to assume that all buildings have known or at least similar orientation or that a single plane will be visible in all views. Nor will it always be possible to obtain photographs containing three vanishing points associated with non-degenerate orthogonal directions. We proceed by making only the following assumptions: firstly that the vertical direction can be identified in all views and secondly that we have a map of the scene.

Whilst a number of previous works have explored the possibility of using a map as an affine view of a scene in combination with one [22] or two [13] perspective views (and additional scene constraints [3]), using a map in combination with many perspective views has not been considered.

We use a map along with the user's prior knowledge of parallelism in order to determine camera orientation and calibration. This allows us to formulate the *uncalibrated* structure from motion problem as a simple *linear* equation without the problem of a possibly degenerate approach to calibration or the need for restrictive assumptions about the scene (such as that reference points are visible in all images). In addition the map provides a strong *global* constraint on structure from motion, allowing high quality reconstruction from a few, noisy feature correspondences.

We describe a complete interactive system for architectural modelling. In comparison with existing systems, our system allows us to build much larger-scale models more quickly.

1.3 Structure of This Paper

This paper is arranged as follows. Section 2 reviews briefly the theory of perspective and affine projection. Section 3 describes how parallelism and map constraints may be used to determine camera calibration and camera registration in the map-based frame. Section 4 explains how these techniques are implemented in a working system. Finally Section 5 presents some experiments to demonstrate the efficacy of these ideas when applied to building large-scale architectural models.

2 Theory and Notation

A 3D point $\mathbf{X}_j = [X \quad Y \quad Z \quad 1]^t$ projects into an image plane according to a general 3×4 projection matrix $\mathbf{P}_i$:

$$\mathbf{x}_{ij} \sim \mathbf{P}_i \mathbf{X}_j \tag{1}$$

where $\mathbf{x}_{ij} = [u \quad v \quad 1]^t$ is an image coordinate and $\sim$ means equality up to scale.

A projection matrix corresponding with a perspective camera may be decomposed as:

$$\mathbf{P}_i = \mathbf{K}_i \,[\mathbf{R}_i \quad -\mathbf{R}_i^t \mathbf{T}_i\,] \tag{2}$$

where $\mathbf{R}_i$ is the 3×3 rotation matrix describing the orientation of the camera and $\mathbf{T}_i$ is the position of the camera. Camera calibration matrix $\mathbf{K}_i$ is of the form:

$$\mathbf{K}_i = \begin{bmatrix} \alpha_u & s & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{3}$$

where α_u and α_v are scale factors, s is skew, and $[\,u_0 \quad v_0 \quad 1]^t$ is the principal point.

The map coordinate of a 3D point is dependent on scene X, Y position but not on Z-axis height. Thus, a map may be modelled as an affine (or orthographic) view of the scene with projection matrix:

$$\mathbf{P}_{\text{map}} \sim \begin{bmatrix} \sigma & 0 & 0 & X_0 \\ 0 & \sigma & 0 & Y_0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{4}$$

where σ is the map's scale and $[\,X_0 \quad Y_0\,]^t$ is the world X, Y coordinate of the map's origin.

3 Map Constraints

3.1 Single View Constraints

The image coordinate $\mathbf{e}_i$ of the vertical vanishing point in image i may be determined from the image of two or more vertical lines. This vanishing point is the projection of the point $[\,0 \quad 0 \quad 1 \quad 0\,]^t$ at infinity and is the epipole corresponding with the affine map camera.

We begin by rectifying our images such that vertical lines in the world map to vertical lines in the image plane. We seek a 3 × 3 homography such that:

$$\mathbf{H}_i\,[\,e_1 \quad e_2 \quad e_3\,]^t = [\,0 \quad 1 \quad 0\,]^t \tag{5}$$

where $\mathbf{e}_i = [\,e_1 \quad e_2 \quad e_3\,]^t$ is the image coordinate of the vanishing point corresponding with the vertical direction in the world and $|\mathbf{e}_i| = 1$. It is convenient to choose $\mathbf{H}_i$ such that $\mathbf{H}_i$ is a rotation matrix and:

$$\mathbf{H}_i\,[\,e_2 \quad -e_1 \quad 0\,]^t = [\,1 \quad 0 \quad 0\,]^t \tag{6}$$

This transformation preserves scale along the line in the image plane that is parallel with the image of the horizon and passes through the point $[\,0 \quad 0 \quad 1\,]^t$. Figure 1 illustrates transformation of an image plane by such a homography.

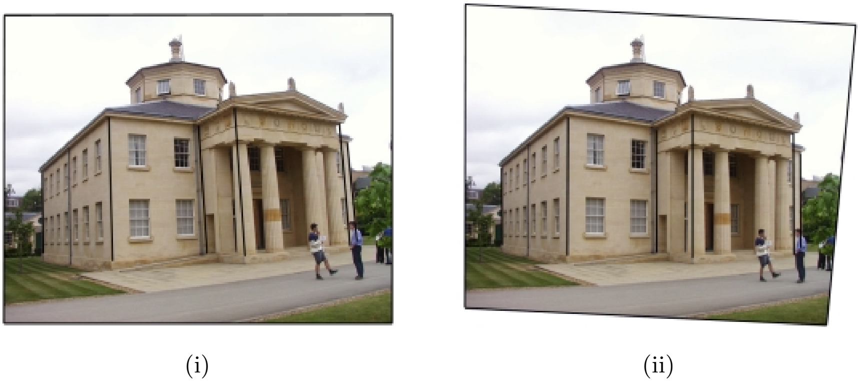

(i) (ii)

Fig. 1. (i) Vertical lines marked in an image. (ii) The image warped by a homography **H** such that vertical lines project to vertical lines in the transformed image plane

3D points $\mathbf{X}_j$ project into our transformed image plane according to the following equation:

$$\mathbf{x}'_{ij} \sim \mathbf{H}_i\mathbf{K}_i\,[\,\mathbf{R}_i \quad -\mathbf{R}_i^t\mathbf{T}_i\,]\,\mathbf{X}_j \sim \hat{\mathbf{P}}_i\mathbf{X}_j \tag{7}$$

where $\mathbf{x}'_{ij} = [\,u' \quad v' \quad 1\,]$ and $\hat{\mathbf{P}}_i$ has the general form:

$$\hat{\mathbf{P}}_i = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \tag{8}$$

In the rectified image, u' coordinates depend only on world X, Y coordinates and are independent of world Z coordinates. Thus $p_{13} = p_{33} = 0$ and transformed image coordinates $\mathbf{X}'$ are related to map X, Y coordinates by the simple 1D projection relationship:

$$\begin{bmatrix} u' \\ 1 \end{bmatrix} \sim \begin{bmatrix} p_{11} & p_{12} & p_{14} \\ p_{31} & p_{32} & p_{34} \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = \mathbf{p}_i \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \tag{9}$$

where $\mathbf{p}_i$ is a 1D projection matrix. Given five or more correspondences between map X, Y coordinates and image u' coordinates (for points or vertical lines), it is possible to solve for all the elements of $\mathbf{p}_i$ as the solution of a linear equation. We can also use as correspondences horizontal vanishing points in the image corresponding with known directions $\mathbf{d}_k$ on the map. Map directions can be estimated from a single line or two or more parallel lines. Thus, we may solve for $\mathbf{p}_i$ given five point correspondences or four point correspondences plus one horizontal vanishing point or three point correspondences plus two horizontal vanishing points.

In a similar manner to the 3 × 4 projection matrix in equation (1), the projection matrix $\mathbf{p}_i$ for a 1D camera may be decomposed as:

$$\mathbf{p}_i = [\,\mathbf{k}_i\mathbf{r}_i \quad -\mathbf{r}_i^t\mathbf{t}_i\,] \tag{10}$$

where $\mathbf{k}_i$ is upper triangular, $\mathbf{r}_i$ is 2 × 2 rotation matrix describing the 1D camera's orientation about the vertical axis, and $\mathbf{t}_i$ is the camera's X, Y map position. Note that there is an ambiguity associated with our solution for the elements of $\mathbf{p}_i$ relating to whether 3D points are in front of or behind the camera. If $\mathbf{r}_i$ is such that X, Y coordinates are behind the 1D camera then we should replace it with $-\mathbf{r}_i$.

The 1D projection matrix $\mathbf{p}_i$ describes vertical axis orientation, map position, and calibration for the 1D camera. Figure (2) illustrates the registration of 1D cameras on the map using five map coordinates, four map coordinates plus one map direction, and three map coordinates plus two map directions.

Given the 1D projection matrix $\mathbf{p}_i$, we may determine $\mathbf{K}_i$ and $\mathbf{R}_i$ for the original view. From (7), and considering only the first 3 × 3 sub matrix:

$$\mathbf{H}_i\mathbf{K}_i\mathbf{R}_i \sim \hat{\mathbf{P}}_i \tag{11}$$

Since $\mathbf{R}_i$ is a rotation matrix, $\mathbf{R}_i\mathbf{R}_i^t = \mathbf{I}$. Thus:

$$\mathbf{H}_i\mathbf{K}_i\mathbf{K}_i^t\mathbf{H}_i^t = \lambda\hat{\mathbf{P}}_i\hat{\mathbf{P}}_i^t \tag{12}$$

where λ is an unkown scale factor. This relationship contains three equations in the known elements of $\hat{\mathbf{P}}_i$ and the unknown elements of $\mathbf{K}_i\mathbf{K}_i^t$ and λ. By assuming that pixels are square ($s = 0$ and $\alpha_u = \alpha_v$) and that the principal point $[\, u_0 \quad v_0 \quad 1\,]^t$ lies at the image centre, we are able to solve this equation for α^2 ($= \alpha_u^2 = \alpha_v^2$). This set of assumptions is at least sufficiently good to allow the approach to succeed for a wide range of cameras. In any case they may be relaxed during the subsequent multi-camera bundle adjustment stage.

Finally we determine $\mathbf{R}_i$. The epipole $\mathbf{e}_i$ is the projection of the point $[0 \quad 0 \quad 1 \quad 0]^t$:

$$\mathbf{e}_i \sim \mathbf{K}_i\mathbf{R}_i\,[0 \quad 0 \quad 1 \quad 0]^t \tag{13}$$

Thus the third column of $\mathbf{R}_i$ is simply $\pm\mathbf{K}_i^{-1}\mathbf{e}_i$. This sign ambiguity arises because the epipole may correspond with the projection of the 'up' or the 'down' direction. We resolve this ambiguity by assuming that photographs are incorporated into our system 'right way up', i.e. the sign of v coordinate of $\mathbf{e_i}$ indicates whether the epipole is 'up' or 'down'. Equation (13) provides two constraints on the three parameters of $\mathbf{R}_i$. Equation (11) allows us to fix the remaining parameter.

3.2 Multiple View Constraints

Using the single view constraints described in the previous section, we can determine camera calibration, orientation, and map X, Y position. However, the Z coordinate of each camera (height) and 3D structure is unknown.

Having determined camera calibration and orientation, we exploit the linear constraint provided by the following equation (as in [6,17]):

$$\mathbf{K}_i^{-1}\mathbf{x}_{ij} \sim [\,\mathbf{R}_i \quad -\mathbf{R}_i^t\mathbf{T}_i\,]\,\mathbf{X}_j \tag{14}$$

This relationship provides two independent linear equations in the elements of unknown structure $\mathbf{X}_j$ and 3D camera positions $\mathbf{T}_i$.

Optionally, we may wish to employ the constraint that some lines have a known direction $\mathbf{d}_k$:

$$\mathbf{d}_k \times \mathbf{L}_l = 0 \tag{15}$$

where × denotes the cross product, $\mathbf{L}_l = \mathbf{X}_t - \mathbf{X}_u$, and $\mathbf{X}_t$ and $\mathbf{X}_u$ are two 3D points connected by a line. We set $\mathbf{d}_0 = [\,0 \quad 0 \quad 1\,]^t$ for the vertical direction and $\mathbf{d}_k = [\,x \quad y \quad 0\,]^t$ for horizontal directions estimated from the map.

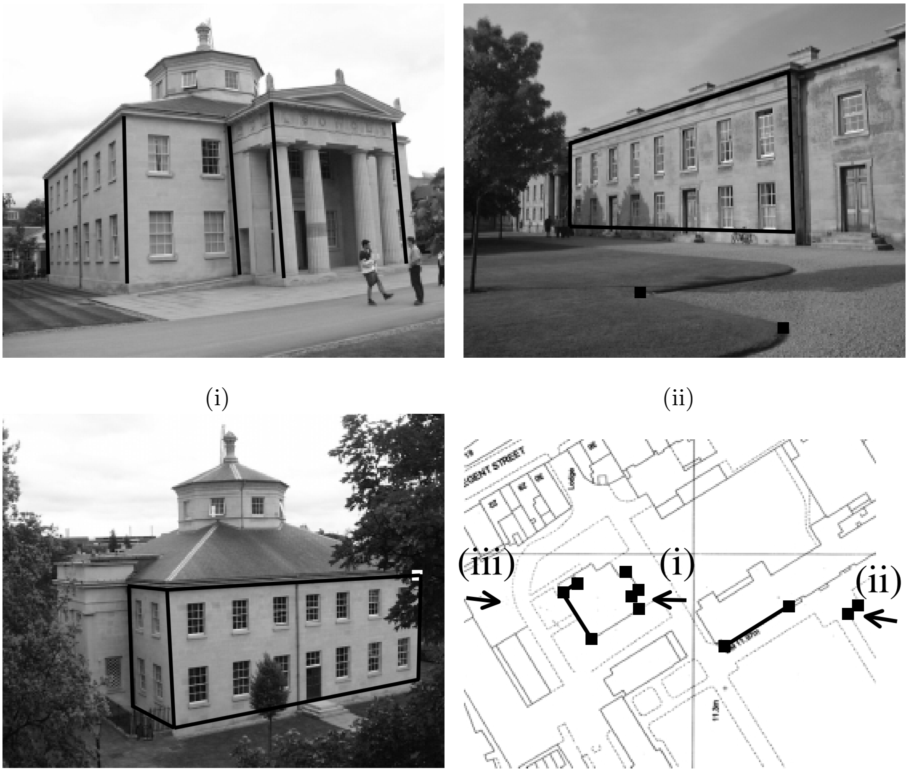
We may assemble all such equations (as 14 and 15) into a matrix equation of the form:

$$\mathbf{AX} = \mathbf{0} \tag{16}$$

where $\mathbf{X}$ comprises all unknown structure $\mathbf{X}_j$ and camera positions $\mathbf{T}_i$.

This equation can be solved easily for hundreds of vertices using the singular value decomposition. For more vertices, we should resort to an appropriate sparse matrix technique.

Note that there is a sign ambiguity associated with the solution for $\mathbf{X}$ in (16). This ambiguity may be resolved by ensuring all points (or at least the majority



(i)

(ii)

(iii)

(iv)

Fig. 2. A view is registered in the map coordinate frame using the vertical vanishing point plus (i) five correspondences between points or vertical lines in the image and points on the map, or (ii) four correspondences plus one horizontal direction, or (iii) three correspondences plus two horizontal directions.

of points in case of noisy data) are reconstructed in front of the cameras in which they are visible. In addition, we must fix the height of one point, e.g. the height of the first camera can be set to 0.

3.3 Optimization

From an initial guess at projection matrices and structure we can optimise camera parameters and structure by bundle adjustment (see Section 4.2).

4 Implementation

4.1 Algorithm

Our approach to modelling is as follows:

1. We transform each image by a homography such that vertical lines in the world project to vertical lines in the image plane
2. For each camera we estimate absolute orientation and camera calibration using the single-view constraints described in Section 3.1.
3. Given camera calibration and orientation and (optionally) extra scene constraints, we compute camera positions and scene structure as the solution of a linear equation as described in Section 3.2.
4. We optimise scene structure and camera parameters using bundle adjustment.

4.2 Bundle Adjustment

We wish to optimise camera parameters and structure subject to the constraint that parallel lines are parallel (both vertical and horizontal) and our knowledge of the affine projection matrix for the map.

We adjust the parameters of projection matrices $\mathbf{P}_i$, structure $\mathbf{X}_j$ and directions $\mathbf{d}_k$ in order to minimise back-projection error ϵ:

$$\epsilon = \sum_{ij} |\mathbf{P}_i \mathbf{X}_j - \mathbf{X}_{ij}|^2 + \sum_{lk} |\mathbf{d}_k \times \mathbf{L}_l|^2 \tag{17}$$

An initial guess at horizontal directions is obtained from the map. The affine projection matrix for the map camera and the vertical direction are fixed.

We have implemented the fast bundle adjustment algorithm in [18]. We extend this algorithm by including extra parameters corresponding with the unknown focal length of each camera and line directions $\mathbf{d}_k$. In addition we provide the facility to incorporate a (fixed) affine camera (the map).

This algorithm allows us to introduce covariance matrices describing the error p.d.f. associated with feature coordinate measurements. In practice, this allows us to account for the fact that map data may be substantially less accurate than image data.

5 Results

5.1 Camera Registration

Figures 3(i, ii, iii) show representative images from a 16-image sequence. This sequence was obtained on level ground using a digital camera mounted on a tripod, which was positioned at regular intervals along a straight line. Using a 1:500 scale Ordnance Survey map, camera registration was determined by the approach described in this paper.

Figure 3(iv) compares estimated camera X, Y positions with ground truth data. Most of the errors associated with estimated camera positions are in the direction parallel to the viewing direction. This is due to the inevitable ambiguity between depth and focal length in views of scenes that do not occupy a substantial range of depth (RMS error associated with estimated focal lengths was

12.7% of the true value). Recovered camera heights are much more consistent because there is no such ambiguity in the estimate (see Figure 3(v)).

Note that in this sequence the sequential approach to camera registration failed after incorporating only the first few images due to the failure of bundle adjustment to converge to the global minimum. Defining only 10 feature correspondences on the map has allowed registration of all images simultaneously and accurately as the solution of a linear equation. The map-based approach obtains a good reconstruction more quickly with fewer feature correspondences - this is important for an interactive system.

5.2 Large-Scale Models

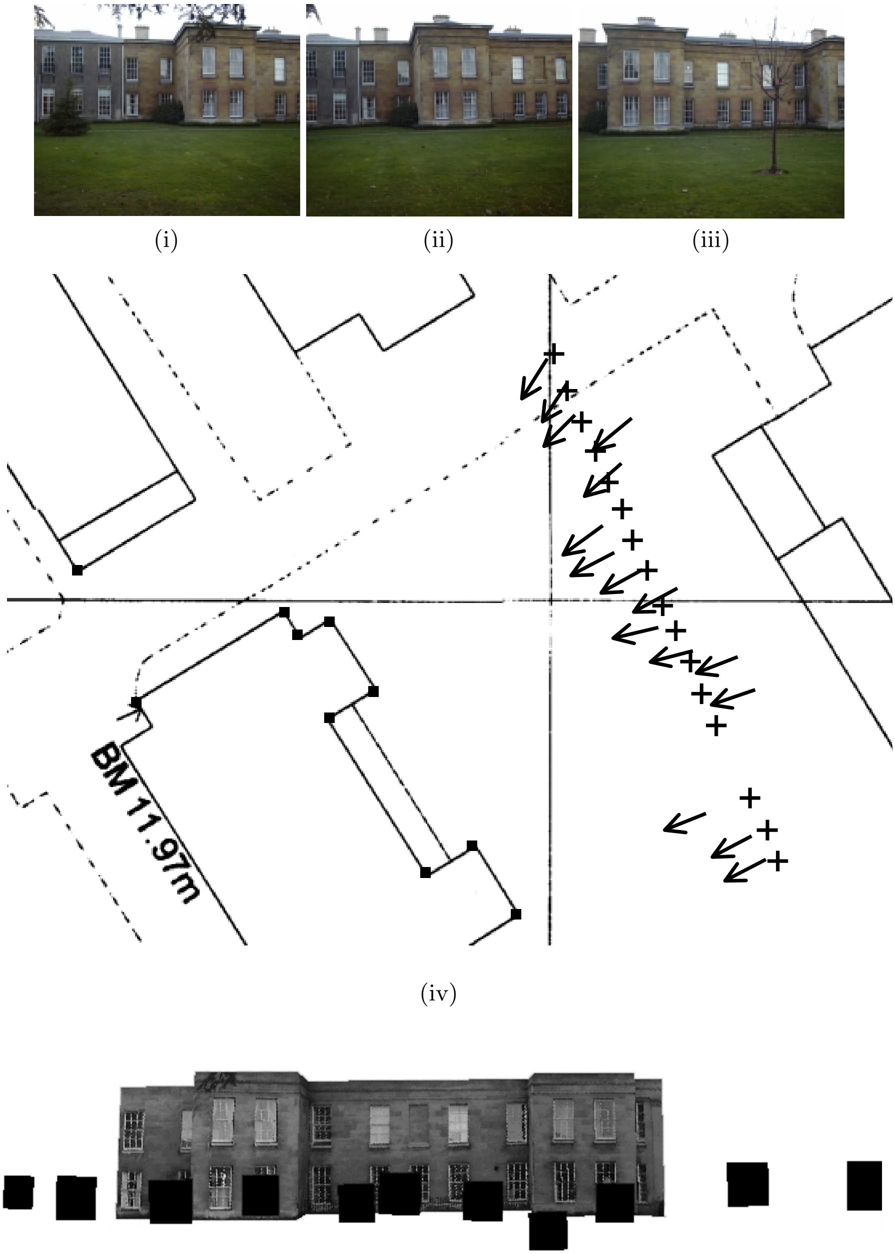
Using the approach described in this paper, we have been able quickly to create large-scale models of a variety of city scenes. Figure 4, for example, shows a view of part of a model of Downing College (before optimisation). Note that since projection matrices are obtained as well as structure, we can also reconstruct features that are not visible on the map.

Compared with existing interactive modelling strategies, the use of the map as an affine view means that more accurate models can be produced more quickly using fewer feature correspondences. Because all camera positions and scene structure are determined simultaneously as the solution of a linear equation, failure of the algorithm is far less common and time-consuming than in systems relying on a sequential DLT plus bundle adjustment approach (like that in [1]). Where the sequential approach fails it is necessary to repeat multiple time-consuming bundle adjustment steps.

5.3 Plane Rectification

Many indoor scenes have planar ceilings and floors, and walls arranged at right angles to each other. This allows us to use the plane rectification technique in [12] to obtain a plan view from photographs of a floor or ceiling such as Figure 5(i). The ceiling plan in Figure 5(ii) is obtained (up to scale) from three rectified photographs without any knowledge of camera focal length. This plan is used like a map view in order to obtain the reconstruction in Figure 5(iii, iv) by the approach described in this paper.

In general modelling indoor scenes is extremely difficult without calibrated wide-angle (or panoramic) cameras. This difficulty arises because only narrow baseline views can be obtained in cramped indoor conditions and degenerate points are common (so that sequential camera registration approaches often fail). By first obtaining a ceiling (or floor) plan we are able to model indoor scenes with ease using a camera with unknown and varying focal length. Our results are comparable with those produced using calibrated panoramas in [17].



(i) (ii) (iii)

(iv)

(v)

Fig. 3. (i, ii, iii) Three representative images from a sequence of 16. (iv) X, Y component of recovered camera positions using the sequential approach compared with ground truth. (v) A synthesised view showing camera image planes at consistent height.

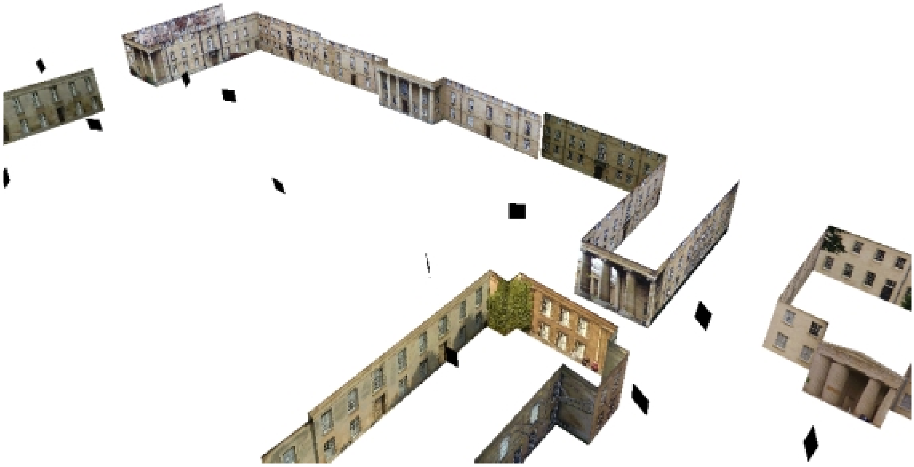


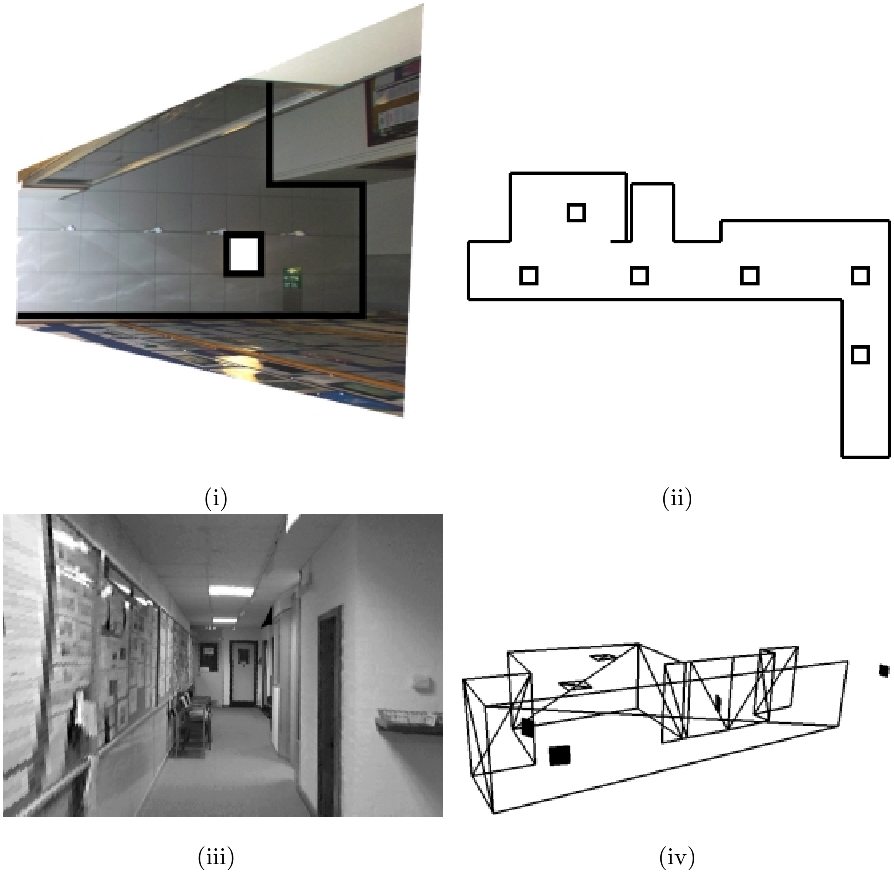
Fig. 4. Part of a large model of Downing College obtained using the approach presented in this paper (before optimisation). This model was reconstructed from 30 photographs obtained from sparse viewpoints using a camera with unknown and varying focal length and a readily available 1:500 scale Ordnance Survey map. Projection matrices were obtained as well as structure allowing reconstruction of points not shown on the map (camera image planes are shown in black).

6 Conclusion

We have developed a practical system for making large-scale architectural models from many uncalibrated images. By using a map along with prior knowledge of which lines are vertical, we have shown that the *uncalibrated* structure from motion problem can be formulated as a simple *linear* equation. In the context of an interactive system, this reconstruction approach succeeds where the sequential approach fails and does not rely on overly restrictive assumptions about the scene.

Although map information may be *locally* much less accurate than image data (e.g. on the scale of a single building), it does provide a strong constraint on *absolute* geometry. Thus, map information may be used in combination with image data interactively to build much larger models than can be obtained using images alone. This approach makes possible to build models of whole city streets rather than simply a few buildings. An additional benefit is that models are registered in an absolute (map) coordinate system.

The principal limitation of all interactive approaches to model building is the amount of time required of the user. However, a number of techniques may be used automatically to improve coarse models produced interactively (e.g. voxel



(i) (ii)

(iii) (iv)

Fig. 5. (i) One of three rectified photographs of the ceiling from which a complete map (ii) was assembled. The map allows recovery of structure and camera position as the solution of a linear equation. (iii) A synthesised novel view. (iv) Recovered structure and camera positions (before optimisation).

carving [15] and template-based model fitting [7]). Critical to the success of these techniques is a good initial guess at camera registration and structure. Present work concerns supplementing fast interactive modelling techniques with automatic ones.

References

1. P. Beardsley, P. Torr, and A. Zisserman. "3D Model Acquisition from Extended Image Sequences." In Proc. *4th European Conference on Computer Vision*, Cambridge (April 1996); LNCS 1065, volume II, pages 683-695, Springer-Verlag, 1996.

2. P. Beardsley, A. Zisserman, and D. Murray. “Sequential Updating of Projective and Affine Structure from Motion.” *International Journal of Computer Vision (23)*, No. 3, Jun-Jul 1997, pages 235-259.
3. D. Bondyfalat, T. Papadopoulo, B. Mourrain. “Using Scene Constraints during the Calibration Procedure.” In Proc. *ICCV’01*, volume II, pages 124-130.
4. B. Caprile and V. Torre. “Using vanishing points for camera calibration.” *International Journal of Computer Vision*, 4(2):127–140, March 1990.
5. R. Cipolla, T. Drummond and D. Robertson. “Camera calibration from vanishing points in images of architectural scenes.” In Proc. *British Machine Vision Conf.*, volume II, pages 382-392, 1999.
6. P. E. Debevec, C.J. Taylor, and J. Malik. “Modelling and Rendering Architecture from Photographs: A Hybrid Geometry- and Image-Based Approach.” In *A CM Computer Graphics (Proceedings SIGGRAPH)*, pages 11-20, 1996.
7. A. Dick, P.Torr, S. Ruffle, R. Cipolla. “Combining Single View Recognition and Multiple View Stereo for Architectural Scenes.” In /em Proc. 8th IEEE International Conference on Computer Vision (ICCV’01), pages 268-274, July 2001.
8. A. W. Fitzgibbon and A. Zisserman. “Automatic Camera Recovery for Closed or Open Image Sequences.” In Proc. *ECCV*, 1998.
9. R. I. Hartley. “Euclidean reconstruction from uncalibrated views.” In J.L. Mundy, A. Zisserman, and D. Forsythe, editors, *Applications of Invariance in Computer Vision*, volume 825 of Lecture notes in Computer Science, pages 237-256, Springer-Verlag, 1994.
10. R. I. Hartley and P. Sturm. “Triangulation.” In *American Image Understanding Workshop*, pages 957-966, 1994.
11. R. I. Hartley. “In defence of the 8-point algorithm.” In Proc. *International Conference on Computer Vision*, pages 1064-1070, 1995.
12. D. Liebowitz and A. Zisserman. “Combining Scene and Auto-calibration Constraints.” In Proc. *ICCV*, volume I, pages 293-300, 1999.
13. N. Navab, Y. Genc, and M. Appel. “Lines in one orthographic and two perspective views.” In Proc. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR’00)*, Hilton Head Island, South Carolina, Vol. 2, 607-616, June 2000.
14. M. Pollefeys, R. Koch, and L. Van Gool. “Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters.” In *Proc. 6th Int. Conf. on Computer Vision*, Mumbai, India, 1998.
15. K. N. Kutulakos and S.M. Seitz. “A theory of shape by space carving.” In *Seventh International Conference on Computer Vision (ICCV’99)*, Greece, September 1999.
16. C. Rother and S. Carlsson: “Linear Multi View Reconstruction and Camera Recovery.” In Proc. *8th International Conference on Computer Vision (ICCV’01)*, July 2001, Vancouver, Canada, pp. 42-49
17. H-Y. Shum, M. Han, and R. Szeliski. “Interactive Construction of 3D Models from Panoramic Mosaics.” In Proc. *IEEE Conf. Computer Vision and Pattern Recognition*, pages 427-433, Santa Barbara, (June) 1998.
18. C. Slama. “Manual of Photogrammetry.” American Society of Photogrammetry, Falls Church, VA, USA, 4th edition, 1980.
19. P. F. Sturm and S. J. Maybank. “A Method for Interactive 3D Reconstruction of Piecewise Planar Objects from Single Images.” In Proc. *British Machine Vision Conference*, volume I, pages 265-274, 1999.
20. I. E. Sutherland. “Three dimensional data input by tablet.” Proc. *IEEE*, Vol 62, No. 4:453-461, April 1974.

21. M. Antone and S. Teller. “Automatic Recovery of Relative Camera Rotations for Urban Scenes.” Proc. *CVPR 2000*, Volume II, June 2000, pp. 282-289.
22. Z. Zhang, P. Anandan, and H.-Y. Shum. “What can be determined from a full and a weak perspective image?” *International Conference on Computer Vision (ICCV’99)*, Corfu, Greece, pages 680-687, September 1999.