MVA2000 IAPR Workshop on Machine Vision Applications, Nov. 28–30, 2000, The University of Tokyo, Japan

11—2

Reconstruction of architectural scenes from uncalibrated photos and maps.

Ignazio Infantino *,Roberto Cipolla †,Antonio Chella ‡

* ‡ Dipartimento di Ingegneria Elettrica, Universita' di Palermo, Italy.
† Department of Engineering, University of Cambridge, UK.

Abstract

In this paper we consider the problem of reconstructing architectural scenes from multiple photographs taken from arbitrarily viewpoints. The original contribution of this work is the use of a map as a source of geometric constraints in order to obtain in a fast and simple way a detailed model of a scene. We suppose images are uncalibrated and have at least one planar structure as a façade for exploiting the planar homography induced between world plane and image to calculate a first estimation of the projection matrix. Estimations are improved by using correspondences between images and map. We show how these simple constraints can be used to calibrate the cameras and recover the projection matrices for each viewpoint. Finally, triangulation is used to recover 3D models of the scene and to visualise new viewpoints. Our approach needs minimal a priori information about the camera being used. A working system has been designed and implemented to allow the user to interactively build a model from uncalibrated images from arbitrary viewpoints and a simple map.

1 Introduction

The aim of the work in this paper is to be able to reconstruct architectural sites from uncalibrated images and using information from the facades and maps of buildings. Many approaches exist to attempt to recover 3D models from calibrated stereo images [15] or uncalibrated extended image sequences [1, 16, 21] by triangulation and exploiting epipolar [14] and trilinear constraints [10]. Other approaches consists of visualisation from image-based representations of a 3D scene. This approach has been successfully used to generate an intermediate viewpoint image given two nearby viewpoints and has the advantage that is does to needs to make explicit a 3D model of the scene [7, 8, 17, 19, 20]. Constructions of 3D model from a collection of panoramic image mosaics and geometrical constraints have also presented [11, 18]. In this paper we adopt simple approach to construct a 3D model by exploiting strong constraints present in the scenes to be modelled[3, 5, 13]. The constraints which can be used are parallelism and orthogonality, leading to simple and geometrically intuitive methods to calibrate the intrinsic and extrinsic parameter of the cameras and to recover Euclidean models of the scene from only two images from arbitrary positions. We propose an extension to [3] dealing with the problem of recovering 3D models from uncalibrated images of architectural scenes viewing façades by exploiting the planar homography induced between world and image and using a simple map.

*Address: Viale delle Scienze, 90100, Palermo, Italy. E-mail: ignazio@csai.unipa.it

†Address: Trumpinghton Street, CB2 1PZ, Cambridge, UK. E-mail: cipolla@eng.cam.ac.uk

‡Address: Viale delle Scienze, 90100, Palermo, Italy. E-mail: chella@unipa.it

2 Modelling system

Our modelling system uses one or more pairs of images of a façade. For each pair, the user indicates line, point or plane correspondences between images and the map. The modelling system attempts to use all possible constraints in a consistent and coherent way. The goal is reached by decomposing the process into several linear steps.
For a single view we perform:

1. Recovering the camera intrinsic parameters and rotation $\mathbf{R}$ from two vanishing points;
2. Recovering camera translation $\mathbf{t}$ from two known points on the map;
3. Rectification of the image based on the homography induced in the image of the planar structure.

In this way we have a first approximation of the projection matrix and the texture maps of planar

structures [12]. These can be directly placed in a 3D model by the map correspondences. In order to have a better estimation of the projection matrix we use two strategies. The first one exploits two images viewing the same planar structure [20] and it deals with new homography estimation by automatically finding corner correspondences. Decomposition of inter-frame homography matrix gives the new estimation of the projection matrices of two views. The second one involves a global refinement using all possible map constraints. It allows reconstruction of points out of the planes used for calibration and rectification. The final steps are:

1. Global optimisation using map constraints (bundle adjustment);
2. Triangulation and raw 3D model;
3. Texture mapping;
4. Export a VRML model.

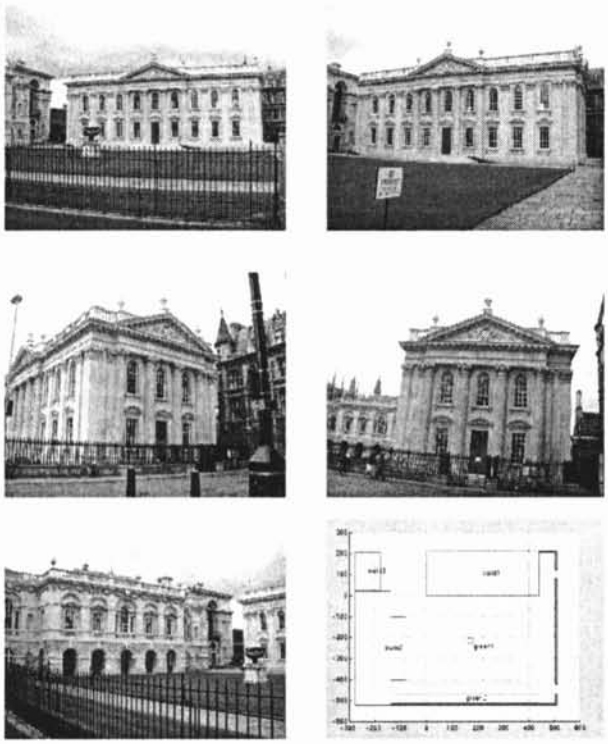

Figure 1: Five of the ten input images of an architectural scene and its map.

3 Calibration and estimation of the projection matrix

For a pin-hole camera, perspective projection from Euclidean 3D space to an image can be represented in homogeneous coordinates by a 3 x 4 camera projection matrix $\mathbf{P}$ [6]:

$$\lambda_i \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \mathbf{P} \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix} \tag{1}$$

The projection matrix has 11 degrees of freedom and can be decompose into a 3 x 3 orientation matrix $\mathbf{R}$ and a 3 x 1 translation vector $\mathbf{t}$ and a 3 x 3 camera calibration matrix, $\mathbf{K}$:

$$\mathbf{P} = \mathbf{K} \, [\, \mathbf{R} \; \mathbf{t} \,] \tag{2}$$

From line correspondences given by the user, we can calculate vanishing points. In the worst case we only have information about two perpendicular direction, but it is enough to estimate the scale factor α of the image plane if we suppose that the principal point coordinates are equal to image centre. Then we can calculate the camera calibration matrix $\mathbf{K}$ and the rotation matrix $\mathbf{R}$ for each image [2]. Moreover if we use two point correspondences between the map and the image (or a known length of a line on image) we also have the translation vector $\mathbf{t}$. From a common planar structure in the pair of images it is possible to automatically improve the estimated matrices above by exploiting the homography[6]. If we suppose points are on plane Z=0, we can write:

$$\begin{aligned} \lambda_1 \; \mathbf{w}_1 &= \mathbf{K}_1 \; [\mathbf{r}_1^1 | \mathbf{r}_1^2 | \mathbf{t}_1] \; \mathbf{X}^{\mathbf{P}} \\ \lambda_2 \; \mathbf{w}_2 &= \mathbf{K}_2 \; [\mathbf{r}_2^1 | \mathbf{r}_2^2 | \mathbf{t}_2] \; \mathbf{X}^{\mathbf{P}} \end{aligned} \tag{3}$$

where $\mathbf{r}_1^1$,$\mathbf{r}_1^2$ are the two first columns of $\mathbf{R}_1$,$\mathbf{r}_2^1$,$\mathbf{r}_2^2$ are columns of $\mathbf{R}_2$, and $\mathbf{X}^{\mathbf{P}}$ is a 3 x 1 vector which denote point is on plane $\Pi(Z = 0)$ in homogeneous coordinates. Let be

$$\begin{aligned} \mathbf{H}_1 &= \mathbf{K}_1 \; [\mathbf{r}_1^1 | \mathbf{r}_1^2 | \mathbf{t}_1] \\ \mathbf{H}_2 &= \mathbf{K}_2 \; [\mathbf{r}_2^1 | \mathbf{r}_2^2 | \mathbf{t}_2] \end{aligned} \tag{4}$$

and

$$\mathbf{H_{21}} = \mathbf{H}_2 \; \mathbf{H}_1^{-1} \tag{5}$$

then

$$\lambda_2 \; \mathbf{w}_2 = \lambda_1 \; \mathbf{H_{21}} \; \mathbf{w}_1 \tag{6}$$

from which we can obtain $\mathbf{w}_2 = [u_2 \; v_2]^T$:

$$\begin{aligned} u_2 &= (\mathbf{h}_1 \; \mathbf{w}_1)/(\mathbf{h}_3 \; \mathbf{w}_1) \\ v_2 &= (\mathbf{h}_2 \; \mathbf{w}_1)/(\mathbf{h}_3 \; \mathbf{w}_1) \end{aligned} \tag{7}$$

where $\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3$ are rows of $\mathbf{H_{21}}$. Fixing for instance a façade, and using a Harris corners detector we can find some correspondences between the

two images. For each detected feature found in the first image we calculate where it is exactly in second image by the homography. Only the stronger matches are selected and we suppose them belonging to viewed plane. In this way we have new estimates of correspondences on second image and we can improve estimate of homography. The decomposition of the homography matrix gives a new estimation of the inter-frame projection matrix.

$$\mathbf{H_{21}} = d_{12} * \mathbf{R_{21}} + \mathbf{t_{21}}\mathbf{n_{21}}^T \tag{8}$$

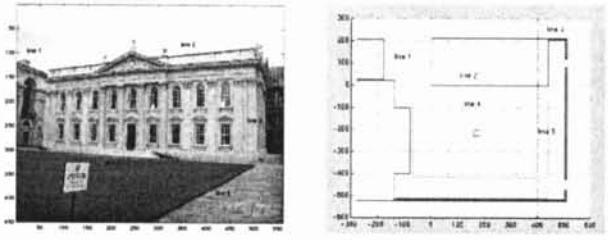

Figure 2: Map constraint example: line to line correspondences.

4 Using map constraints

The map gives us important constraints to improve the estimation of the 3D model. We use the following constraints between geometric entities of image and map [19, 20]:

1. point to point correspondences
2. point to line correspondences
3. line to line correspondences

For each image the user give some correspondences to a map and the system tries to improve the projection matrix estimation. Point to point correspondences are strong constraints because are directly related to the projection matrix:

$$\lambda_{\mathbf{k}}\mathbf{w_k} = \mathbf{P}\mathbf{X_k} \tag{9}$$

Point to line correspondence is expressed by equation:

$$\mathbf{L}\mathbf{P}^{-1}_{\mathbf{4x4}} \begin{bmatrix} \lambda \mathbf{w} \\ 1 \end{bmatrix} = \mathbf{0_{3x1}} \tag{10}$$

where $\mathbf{L}$ is a 3x4 matrix of coefficients of 3 dimensional line equations and $\mathbf{P}^{-1}_{\mathbf{4x4}}$ is inverse matrix of projection matrix $\mathbf{P}$ with adding the last row [0 0 0 1]. 3d line trough points $\mathbf{X_2}$ and $\mathbf{X_1}$ can be expressed by:

$$(\mathbf{X_2} - \mathbf{X_1}) \wedge \mathbf{X} + (\mathbf{X_2} \wedge \mathbf{X_1}) = 0 \tag{11}$$

If we denote with $a = x_2 - x_1$, $b = z_2 - z_1$, $c = y_2 - y_1$, $d_1 = z_1y_2 - y_1z_2$,$d_2 = -z_1x_2 + x_1z_2$, $d_3 = y_1x_2 - x_1y_2$, line is described by a pair of equations from

$$\begin{aligned} by - cz + d_1 &= 0 \\ az - bx + d_2 &= 0 \\ cx - ay + d_3 &= 0 \end{aligned} \tag{12}$$

From equations system (10) we can eliminate λ and use 2 constraints.

Line to line correspondences are expressed as:

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \mathbf{w_1} \wedge \mathbf{w_2} = (\mathbf{P}\mathbf{X_3}) \wedge (\mathbf{P}\mathbf{X_4}) \tag{13}$$

where a, b, c are coefficients of line equation $au + bv + c = 0$ on image passing trough generic points $\mathbf{w_1}$ and $\mathbf{w_2}$, and $\mathbf{X_3}$, $\mathbf{X_4}$ are two generic points of line map.

All constraints given by the user are used to define a minimisation problem in order to obtain new estimates of the projection matrices. Decomposing new matrices by QR decomposition we have also new estimates of camera calibration matrix $\mathbf{K}$, rotation matrix $\mathbf{R}$, translation vector $\mathbf{t}$ and camera position $\mathbf{X_c}$ of each image. In the process used to estimate better projection matrices we introduce also point to point correspondences which arise from the vanishing points calculated. In this way we have always at least two point-to-point correspondences, and we can obtain a reconstruction even if the user doesn't introduce other correspondences. The obtained 3D structure is rendered afterwards using a texture mapping procedure and the final model is stored in standard VRML 1.0 format. In order to preserve the information given by primitive segments, we use a Delaunay triangulation on 2D texture (that is a image used for reconstruction).

5 Conclusion

The techniques presented have been successfully used to interactively build models of architectural scenes from pairs of uncalibrated photographs. Using only information from planar structure such as façades and a simple map, we can recover precise projection matrices with only a few point correspondences.

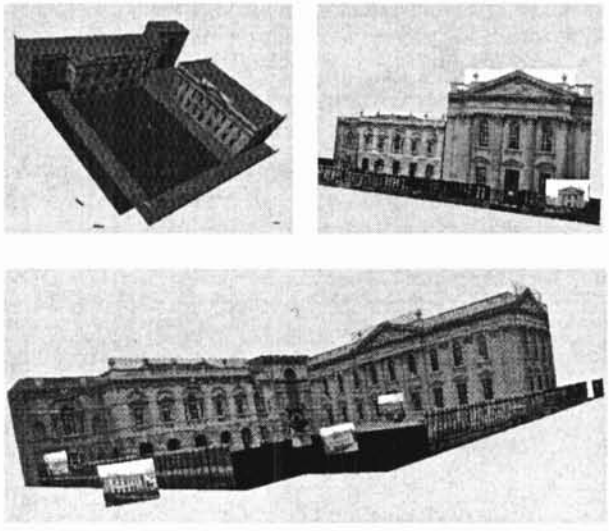

Figure 3: Some views of the final VRML model reconstruction

References

[1] P. Beardsley, P. Torr and A. Zisserman. 3D model acquisition from extended image sequences. In *Proc. 4th European Conf. on Computer Vision*, Cambridge (April 96); LNCS 1065, vol. II, pp. 683-695, Springer-Verlag, 1996.

[2] B. Caprile and V. Torre. Using vanishing points for camera calibration. *IJCV*, pp. 127-140, 1990.

[3] R. Cipolla, T. Drummond and D. Robertson. Camera calibration from vanishing points in images of architectural scenes. *n Proc. British Machine Vision Conference*, Nottingham, volume 2, pp. 382-391, 1999.

[4] A. Criminisi, I. Reid, and A. Zisserman. Single View Metrology. In *Proc. 7th International Conference on Computer Vision*, pp.434-442, 1999.

[5] P.E. Debevec, C.J. Taylor, and J. Malik. Modelling and rendering architecture from photographs: a hybrid geometry-and image-base approach. *In ACM Computer Graphics (proceedings SIGGRAPH)*, pp. 11-20, 1996.

[6] O. Faugeras and F. Lustman. Motion and structure from motion in a piecewise planar environment. *Int. Journal of Pattern Recognition and Artificial Intelligence*, 2(3):485-508, 1988.

[7] O. Faugeras, S. Laveau, L. Robert, G. Csurka, and C. Zeller. 3-D reconstruction of urban scenes from sequences of images. *Computer Vision and Image Understanding*, n.69, vol.3, pp: 292-309, 1998.

[8] S.J. Gortler, R. Grzeszczuk, R. Szeliski, and M.F. Cohen. The Lumigraph. In*ACM Computer Graphics (Proc. SIGGRAPH)*, pp: 31-42, 1996.

[9] R. I. Hartley. Estimation of relative camera positions for uncalibrated cameras. *European Conf. on Computer Vision*, pp. 579-587, S. Margherita Ligure, Italy, May 1992.

[10] R. I. Hartley. Lines and points in three views and the trifocal tensor. *International Journal of Computer Vision*, 22(2): 125-140, 1996.

[11] S.B. Kang and R. Szeliski. 3-D scene data recovery using omni-directional multibaseline stereo. In *CVPR'96*, pp. 364-370, June 1997.

[12] D. Liebowitz and A. Zisserman. Metric Rectification for Perspective Images of Planes. in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 482-488, 1998.

[13] D. Liebowitz,A. Criminisi and A. Zisserman. Creating architectural models from images. *Eurographics'99*, vol.18, n.3, 1999.

[14] Q.-T. Luong and T. Viéville. Canonical Representations for the Geometries of Multiple Projective Views. *Computer Vision and Image Understanding*, 64(2): 193-229, 1996.

[15] P.J. Narayanan,P.W. Rander and T. Kanade. Constructing virtual worlds using dense stereo. In *Proc. of Sixth IEEE Intl. Conf. on Computer Vision*, Bombay (India), pp. 3-10, January 1998.

[16] M. Pollefeyes,R. Koch and L. Van Gool. Self calibration and metric reconstruction inspite of varying an unknown internal camera parameters. In *Proc. of Sixth IEEE Intl. Conf. on Computer Vision*, Bombay (India), pp. 90-95, January 1998.

[17] S.M. Seitz, and C.R. Dyer. Toward image-based scene representation using view morphing. In *Proc. of Intl. Conf. IEEE Conf. on Pattern Recognition*, Vienna (Austria), January 1996.

[18] H-Y Shum, M. Han and R. Szeliski. Interactive construction of 3-D models from panoramic mosaics. *In Proc IEEE Conf. On Computer Vision and Pattern Recognition*, pag. 427-433, Santa Barbara (USA), June 1998.

[19] R. Szeliski and S. Heung-Yeung. Creating full view panoramic image mosaics and environment maps. *In SIGGRAPH*, 1997.

[20] R. Szeliski and P. Torr. Geometrically constrained structure from motion: Points on planes. *In European Workshop on 3D Structure from Multiple Images of Large-Scale Environments (SMILE)*, pp.171-186, June 1998.

[21] C. Tomasi,and T. Kanade. Shape and motion from image streams under orthography: a factorisation method. *International Journal of Computer Vision*, 9(2):137-154, 1990.