Electronic Letters on Computer Vision and Image Analysis 5(4):75-82, 2006

Architectural Scene Reconstruction from Single or Multiple Uncalibrated Images

Huei-Yung Lin*, Syuan-Liang Chen+ and Jen-Hung Lin $

* *Department of Electrical Engineering, National Chung Cheng University, Chia-Yi 621, Taiwan*
+ *NewSoft Technology Corporation, Taipei, Taiwan*
$ *Machvision Inc., Hsinchu 30076, Taiwan*

Received 25 August 2006; accepted 16 November 2006

Abstract

In this paper we present a system for the reconstruction of 3D models of architectural scenes from single or multiple uncalibrated images. The partial 3D model of a building is recovered from a single image using geometric constraints such as parallelism and orthogonality, which are likely to be found in most architectural scenes. The approximate corner positions of a building are selected interactively by a user and then further refined automatically using Hough transform. The relative depths of the corner points are calculated according to the perspective projection model. Partial 3D models recovered from different viewpoints are registered to a common coordinate system for integration. The 3D model registration process is carried out using modified ICP (iterative closest point) algorithm with the initial parameters provided by geometric constraints of the building. The integrated 3D model is then fitted with piecewise planar surfaces to generate a more geometrically consistent model. The acquired images are finally mapped onto the surface of the reconstructed 3D model to create a photo-realistic model. A working system which allows a user to interactively build a 3D model of an architectural scene from single or multiple images has been proposed and implemented.

Key Words: 3D Model Reconstruction, Range Image, Range Data Registration.

1 Introduction

3D reconstruction of real scenes is one of the most challenging tasks in computer vision [7]. In this work we have focused on the reconstruction of 3D models of architectural scenes. One major difference between the 3D reconstruction of architectural scenes and general objects is that the former contains easily detectable man-made features such as parallel lines, orthogonal lines, corners, etc. These features are important cues for finding the 3D structure of a building. Most research on architectural scene reconstruction in the photogrammetry community has concentrated on 3D reconstruction from aerial images [8, 2]. Due to long-range photography, aerial images are usually modeled as orthographic projection. Although the orthographic projection model is easier for aerial images, one major drawback is that most of the 3D reconstruction of architectural scenes can only be done on the roofs of the buildings. On the other hand, the perspective projection model is usually

Correspondence to: <lin@ee.ccu.edu.tw>

Recommended for acceptance by <E. Marti>
ELCVIA ISSN:1577-5097
Published by Computer Vision Center / Universitat Autònoma de Barcelona, Barcelona, Spain

needed for close-range photography, which is capable of reconstructing the complete (360 degrees) 3D model of an architectural scene.

3D models of architectural scenes have important application areas such as virtual reality (VR) and augmented reality (AR). Both applications require photo-realistic 3D models as input. A photo-realistic model of a building consists not only the 3D shape of the building (geometric information) but also the image texture on the outer visible surface of the building (photometric information). The geometric and photometric information can be acquired either by range data and intensity images, or by the intensity images recorded by a camera. Allen *et al* [1, 9] created 3D models of historic sites using both range and image data. They first built the 3D models from range data using a volumetric set intersection method. The photometric information was then mapped onto those models by registering features from both the 3D and 2D data sets. To accurately register the range and intensity data, and reduce the overall complexity of the models, they developed range data segmentation algorithms to identify planar regions and determine linear features from planar intersections. Dick *et al* [5] recovered 3D models from uncalibrated images of architectural scenes. They proposed a method which exploited the rigidity constraints usually seen in the indoor and outdoor architectural scenes such as parallelism and orthogonality. These constraints were then used to calibrate the intrinsic and extrinsic parameters of the cameras through projection matrix using vanishing points [3]. The Euclidean models of the scene were reconstructed from two images from arbitrary viewpoints.

In this work, we develop a system for 3D model reconstruction of architectural scenes from one or more uncalibrated images. The input images can be taken from off-the-shelf digital cameras and the camera parameters for 3D reconstruction are estimated from the structure of the architectural scene. The feature points (such as corners) in the images are selected by a user interactively through a graphical user interface. The selected image points are then refined automatically using Hough transform to obtain more accurate positions in subpixel resolution. For a given set of corner points, various constraints such as parallelism, orthogonality, coplanarity are enforced to create a 3D model of the building. Partial 3D models reconstructed from different viewpoints are then registered to a common coordinate system to create a complete 3D model. The texture information is finally mapped onto the building to create a photo-realistic 3D model.

2 Camera Model and Parameter Estimation

The most commonly used camera model is the pinhole camera model. In this model the projection from a point (X_i, Y_i, Z_i) in Euclidean 3-space to a point (x_i, y_i) in the image plane can be represented in homogeneous coordinates by

$$s \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{11} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{bmatrix} \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix} \tag{1}$$

where s is an arbitrary scale factor, and the 3×4 matrix

$$\mathbf{M} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{11} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{bmatrix} \tag{2}$$

is the perspective projection matrix of the camera. The perspective projection matrix can be further decomposed into the intrinsic camera parameters and the relative pose of the camera:

$$\mathbf{M} = \mathbf{K}[\mathbf{R}\ \mathbf{t}] \tag{3}$$

The 3×3 matrix $\mathbf{R}$ and 3×1 vector $\mathbf{t}$ are the relative orientation and translation with respect to the world coordinate system, respectively. The intrinsic parameter matrix $\mathbf{K}$ of the camera is a 3×3 matrix and usually

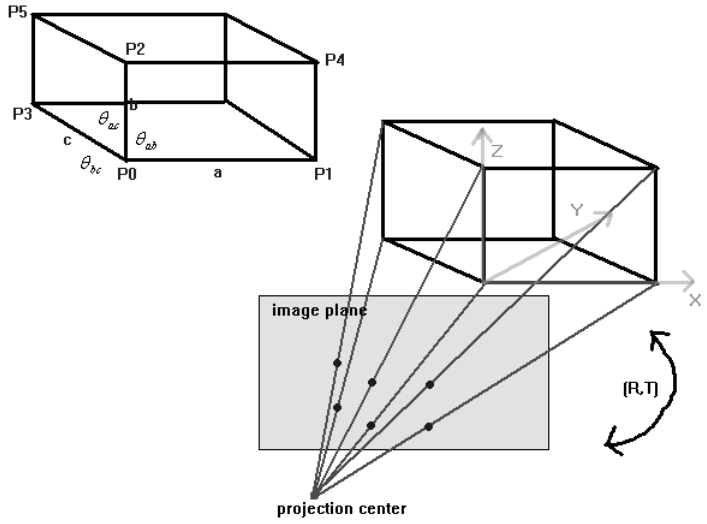

Figure 1: Parallelepiped and a pinhole camera model

modeled as

$$\mathbf{K} = \begin{bmatrix} f_x & \gamma & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{4}$$

where (u_0, v_0) is the principal point (the intersection of the optical axis with the image plane), γ is a skew parameter related to the characteristic of the CCD array, and f_x and f_y are scale factors. Thus, Eq. (1) can be rewritten as

$$s\mathbf{p} = \mathbf{K}[\mathbf{R}\ \mathbf{t}]\mathbf{P} \tag{5}$$

where $\mathbf{P}$ is a 3D point and $\mathbf{p}$ is the corresponding image point (in homogeneous coordinates).

The correctness of reconstructed 3D model depends on the accuracy of camera parameters. Classical camera calibration methods [10] rely on fixed calibration patterns. In this work, the primary goal is to reconstruct 3D models of architectural scenes. Since the man-made structures usually contain parallelepipeds, parallelism and orthogonality will be used for camera parameter estimation. Consider a parallelepiped shown in Fig. 1, which resembles the visible surface of a building. Two planes $\mathbf{P}_0\mathbf{P}_1\mathbf{P}_4\mathbf{P}_2$ and $\mathbf{P}_0\mathbf{P}_2\mathbf{P}_5\mathbf{P}_3$ are determined by the six points, $\mathbf{P}_0$, $\mathbf{P}_1$, $\mathbf{P}_2$, $\mathbf{P}_3$, $\mathbf{P}_4$ and $\mathbf{P}_5$. Assume the corresponding image points are $\mathbf{p}_0$, $\mathbf{p}_1$, $\mathbf{p}_2$, $\mathbf{p}_3$, $\mathbf{p}_4$, $\mathbf{p}_5$, then we have

$$s_i\mathbf{p}_i = \mathbf{K}[\mathbf{R}\mathbf{P}_i + \mathbf{t}] \tag{6}$$

for $i = 0, 1, \cdots, 5$ by Eq. (5).

As shown in [4], for a given parallelepiped in 3D space, if the three angles between its adjacent edges, θ_{ab}, θ_{bc}, θ_{ca}, and the image points of the six points of its two adjacent faces are available, then the pose of the parallelepiped, intrinsic parameters of the camera and the size of the parallelepiped can be determined by solving polynomial equations of at most fourth degree. For a special case that θ_{ab}, θ_{bc} and θ_{ca} are right angles, the equation can be further simplified to a linear system. Thus, the camera parameters can be found if the corner points of a building are identified. Furthermore, the focal length of the camera can be estimated and used for 3D model reconstruction in the next section.

3 Three-Dimensional Model Reconstruction

3.1 Reconstruction Algorithm

Chen *et al* [4] showed that for any given parallelogram in 3D space with known image coordinates of four corner points, the relative depths of the four corner points can be determined. If we consider the four points $\mathbf{P}_0$, $\mathbf{P}_1$, $\mathbf{P}_2$ and $\mathbf{P}_3$, which forms a parallelogram as shown in Fig. 1, then $\mathbf{P}_0 + \mathbf{P}_4 = \mathbf{P}_1 + \mathbf{P}_2$ by the property of parallelogram in any world coordinate. Thus, we have

$$s_0 \begin{bmatrix} u_0 \\ v_0 \\ 1 \end{bmatrix} = A \begin{bmatrix} s_1 \\ s_4 \\ s_2 \end{bmatrix} \tag{7}$$

where A is given by

$$\begin{bmatrix} u_1 & -u_4 & u_2 \\ v_1 & -v_4 & v_2 \\ 1 & -1 & 1 \end{bmatrix} \tag{8}$$

with the corresponding image points $\mathbf{p}_i = (u_i, v_i)$ for $i = 0, 1, 2, 4$. If the three points $\mathbf{p}_1$, $\mathbf{p}_2$, $\mathbf{p}_4$ are not collinear, then A is nonsingular. Thus, the relative depths, s_1/s_0, s_4/s_0, s_2/s_0, of the 3D points are given by

$$\begin{bmatrix} s_1/s_0 \\ s_4/s_0 \\ s_2/s_0 \end{bmatrix} = A^{-1} \begin{bmatrix} u_0 \\ v_0 \\ 1 \end{bmatrix} \tag{9}$$

if the corresponding image points are not collinear.

The above algorithm calculates the relative depths of the object points, i.e., s_1/s_0, s_4/s_0, s_2/s_0, from their corresponding image points $\mathbf{p}_1$, $\mathbf{p}_4$, $\mathbf{p}_2$ with respect to $\mathbf{p}_0$. Only the depth information is not sufficient for shape recovery since the 3D model of the object contains x and y direction information as well. Thus, we use the known focus length of the camera (obtained from camera parameter estimation described in the previous section) to constraint the relative positions of the 3D points. That is, the focus length of the camera is used as a factor to scale the displacement in the x and y directions with respect to the depth of the 3D point.

For example, suppose $\mathbf{P}_0$ and $\mathbf{P}_1$ are two 3D points, then we have

$$\mathbf{P}_0 \sim s_0 \mathbf{p}_0 \tag{10}$$

$$\mathbf{P}_1 \sim s_1 \mathbf{p}_1 \tag{11}$$

by the perspective projection. The ratio of $\mathbf{P}_1$ to $\mathbf{P}_0$ is given as

$$\frac{\mathbf{P}_1}{\mathbf{P}_0} \approx \frac{s_1 \mathbf{P}_1}{s_0 \mathbf{P}_0} = \frac{s_1 [u_1 \; v_1 \; 1]^T}{s_0 [u_0 \; v_0 \; 1]^T} \tag{12}$$

In the above equation, the z direction information is lost in the image coordinate system. To recover the 3D shape of the object, the z coordinate of $\mathbf{P}_0$ is fixed as the focal length of the camera (in pixel), i.e., $z = f$ or $\mathbf{P}_0 = [u_0 \; v_0 \; f]^T$. Now, if $\mathbf{P}_0$ is used as a reference point as shown in Fig. 1, then we have

$$\mathbf{P}_i = \frac{s_i}{s_0} \begin{bmatrix} u_i \\ v_i \\ f \end{bmatrix} \tag{13}$$

for $i = 1, 2, 4$.

3.2 Registration and Pose Estimation

The goal of model registration is to combine two or more partial 3D models acquired from different viewpoints to a complete 3D models. Usually the registration or pose estimation involve finding the rotation matrix and translation vector for the transformation between two different coordinate systems. For any given two partial 3D models of an object, the overlapping parts are used to identify the corresponding 3D points for the two models. The corresponding 3D points are then used to find the rotation matrix and translation vector.

Suppose there are two sets of 3D points to be registered. More precisely, if we want to find the rotation matrix and translation vector for the data sets, $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n\}$ and $\{\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_n\}$, where $\mathbf{x}_i$ and $\mathbf{y}_i$ are the corresponding points, for $i = 1, ..., n$. Then the relationship between $\mathbf{x}_i$ and $\mathbf{y}_i$ can be written as

$$\mathbf{y}_i = \mathbf{R}\mathbf{x}_i + \mathbf{t} \tag{14}$$

Let the correlation matrix for the two data sets be

$$C = \sum_{i=1}^{n} w_i \mathbf{x}_i \mathbf{y}_i^T \tag{15}$$

To find the rotation matrix $\mathbf{R}$, singular value decomposition technique is used to rewrite $\mathbf{C}$ as $\mathbf{C} = \mathbf{U}\mathbf{D}\mathbf{V}^T$ by Eq. (15). Since the rotation matrix must satisfy $\mathbf{R}\mathbf{R}^T = \mathbf{I}$, we have

$$\mathbf{R} = \mathbf{U} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \det(\mathbf{U}\mathbf{V}^T) \end{bmatrix} \mathbf{V}^T \tag{16}$$

by Eq. (15). Finally, the translation vector $\mathbf{t}$ can be calculated by

$$\mathbf{t} = \bar{\mathbf{y}} - \mathbf{R}\bar{\mathbf{x}} \tag{17}$$

where $\bar{\mathbf{x}}$ and $\bar{\mathbf{y}}$ are the centroids of the two data sets, respectively [6].

3.3 Model Optimization

In this work, the 3D model is reconstructed using parallelism and orthogonality of an architectural scene. Thus, selecting feature points (edge points) which form parallel or perpendicular lines is an important issue. In the implementation, a graphical user interface is provided such that the user can manually select the corner points. Since the selected points are not always accurate enough (for example, subpixel resolution is not possible and the selection can be affected by the radial distortion of the camera), an automatic corner detection algorithm is applied on the neighborhood of the selected points. Hough transform is then carried out to find more accurate positions in subpixel resolution. For the images captured with short focal length, the lens radial distortion parameter is approximated by line fitting and then used for image distortion correction.

To use parallelism, orthogonality constraints and force four corner points to be coplanar to create an optimized 3D model, we first find the plane equations for the coplanar points, say A, B, C, D, using least squared fitting. Then the four points are projected on the plane as points A', B', C', D'. Parallelism and orthogonality constraints are applied on the above four points to obtain the points A'', B'', C'', D'' which form a rectangle. After the coplanar four points are determined, similar arguments apply to the planar surfaces which are parallel and orthogonal to each other in order to satisfy the geometric constraints.

If the roof of a building contains different geometric primitives rather than rectangles (e.g., triangles or trapezoids), parallelism and orthogonality constraints cannot be directly applied. In this case, a 3D model of the lower part of the building is first reconstructed as a "base-model", which typically contains two perpendicular planar surfaces. The roof (i.e., upper part) of the building is then reconstructed using coplanarity, parallelism and equidistance constraints on the image points corresponding to the roof and the 3D points on the base-model. Commonly used equidistance constraints include the same length of the left and right sides of a triangle, and the same length of the diagonals of a trapezoid.

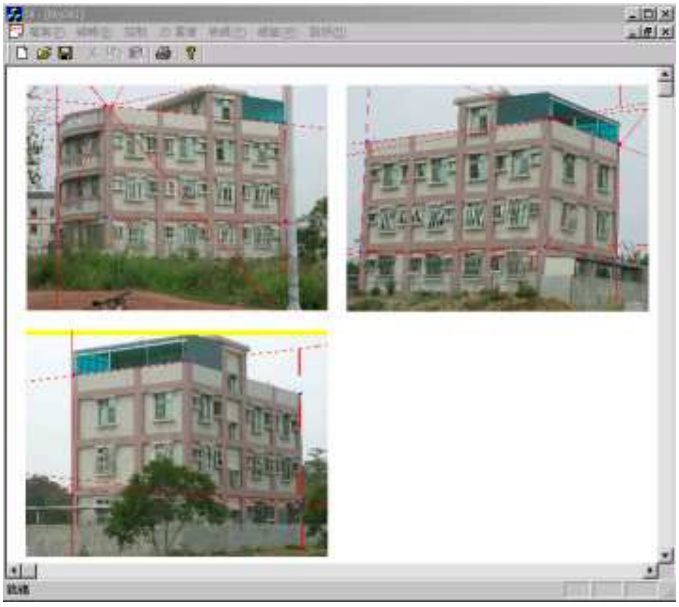

Figure 2: Graphics user interface

4 Experimental Results

The described algorithms are tested on a number of objects for the indoor environment and outdoor architectural scenes. As shown in Fig. 2, a graphics user interface is developed to assist users to select approximate corner points interactively for 3D model reconstruction. The first experiment is the 3D reconstruction of a building. Fig. 2 shows the three images taken from different viewpoints. The images are used to create the partial 3D shapes of the object individually. For each image, the corner points are selected manually by a user and the positions are automatically refined by Hough transform. The reconstructed partial 3D models with texture information using the acquired images are shown in Fig. 3. The complete 3D model after registration and integration is shown in Fig. 4. Although plane fitting has been done on the object surface, rendering with triangular mesh still cause some visual distortions.

For the structures of the objects containing non-rectangular surface patches, camera parameters and the "base-models" are first obtained from the lower part of the objects. Additional image points associated with the upper part of the object are then added with coplanarity and equidistance constraints. Fig. 6 shows the reconstructed partial 3D models from the corresponding single input images shown in Fig. 5. We have tested the proposed 3D model reconstruction approach on six outdoor architectural scenes and four indoor objects. The small scale objects in the laboratory environment usually give better reconstruction results mainly because of the controlled illumination conditions and the larger focal length used for image acquisition. For the outdoor building reconstruction, careful selections of the initial corner points are mandatory since the images might contain more complicate background scenes. Furthermore, the lens distortion has to be modeled for even close-range photography with short focal length. Since the evaluation of the final reconstruction result is usually based on the texture information of the object, novel views are best synthesized on the viewpoints closer to the original acquired images.

5 Conclusion and Future Research

In this paper, we have presented a 3D model reconstruction system for architectural scenes using one or more uncalibrated images. The images can be taken from off-the-shelf digital cameras. The feature points for 3D

Figure 3: Partial 3D shapes of the first experiment

Figure 4: Complete 3D model of the first experiment

model reconstruction are selected by a user interactively through a graphical user interface. For a given set of corner points from one viewpoint, parallelism, orthogonality and coplanarity constraints are applied to obtain the partial 3D shape of the building. The complete (360 degrees) 3D model of the building is obtained by registering the partial 3D models to a common coordinate system. Future research will focus on using additional geometric constraints to create more detailed 3D models of architectural scenes. Whenever it is possible, texture mapping which contains occluding objects such as trees will be avoided by using the images captured from different viewpoints.

Acknowledgments

The support of this work in part by the National Science Council of Taiwan, R.O.C. under Grant NSC-93-2218-E-194-024 is gratefully acknowledged.

References

[1] P.K. Allen, I. Stamos, A. Troccoli, B. Smith, M. Leordeanu, Y.C. Hsu, "3D Modeling of Historic Sites Using Range and Image Data," *IEEE International Conference on Robotics and Automation*, pp. 145-150, 2003

Figure 5: Input images of the second experiment

Figure 6: Partial 3D shapes of the second experiment

[2] C. Baillard and A. Zisserman, "Automatic Reconstruction of Piecewise Planar Models from Multiple Views," *IEEE Computer Vision and Pattern Recognition*, pp. II: 559-565, 1999.

[3] B. Caprile and V. Torre, "Using Vanishing Points for Camera Calibration," *International Journal of Computer Vision*, pp. 127-140, March, 1990.

[4] C. Chen, C. Yu, and Y. Hung, "New Calibration-free Approach for Augmented Reality based on Parameterized Cuboid Structure," *International Conference on Computer Vision*, pp. 30-37, 1999.

[5] A.R. Dick, P.H.S. Torr, and R. Cipolla, "Automatic 3D Modelling of Architecture," *British Machine Vision Conference*, pp. 372-381, 2000.

[6] K. Kanatani, *Geometric Computation for Machine Vision*, Oxford Science Publications, 1993.

[7] E. Trucco and A. Verri, *Introductory Techniques for 3-D Computer Vision*, Prentice Hall, 1998.

[8] T. Moons, D. Frere, J. Vandekerckhove, L.J.V. Gool, "Automatic modelling and 3D reconstruction of urban house roofs from high resolution aerial imagery"," *European Conference on Computer Vision*, pp. I: 410-425, 1998.

[9] I. Stamos and P.K. Allen, "3-D Model Construction using Range and Image Data," *IEEE Computer Vision and Pattern Recognition*, pp. I: 531-536, 2000.

[10] R.Y. Tsai, "A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses," *IEEE Trans. Robotics and Automation*, 3(4), pp. 323-344, 1987.