

3D reconstruction based on spatial vanishing information

Yuan Shu (舒 远) and Zheng Tan (谈 正)

School of Electronic Information and Engineering, Xi'an Jiaotong University, Xi'an 710049

Received May 21, 2004

An approach for the three-dimensional (3D) reconstruction of architectural scenes from two un-calibrated images is described in this paper. From two views of one architectural structure, three pairs of corresponding vanishing points of three major mutual orthogonal directions can be extracted. The simple but powerful constraints of parallelism and orthogonal lines in architectural scenes can be used to calibrate the cameras and to recover the 3D information of the structure. This approach is applied to the real images of architectural scenes, and a 3D model of a building in virtual reality modelling language (VRML) format is presented which illustrates the method with successful performance.

OCIS codes: 100.6890, 120.0120, 120.6650, 100.0100.

Considerable efforts have been made to recover photorealistic three-dimensional (3D) models of the real world. The most common geometric approach is to recover 3D models from calibrated or uncalibrated two-dimensional (2D) images^[1,2]. Camera calibration is an indispensable step to obtain 3D geometric information from 2D images, and many camera calibration methods have been proposed in different ways of recent years^[3,4]. After the camera has been calibrated, 3D structure can be recovered from multiple 2D images using corresponding image points.

A new technique of recovering 3D models from two uncalibrated images of architectural building using spatial vanishing information is presented in this paper. We show how these constraints can be used to calibrate the camera and to recover the projection matrices for each viewpoint. The projection matrices are used to recover partial 3D models of the scene and these can be used to visualize new viewpoints. Our approach does not need any a priori information about the cameras being used, and it has no any constraints on the camera motion. It is suitable for the hand held digital cameras that the camera motion cannot be controlled.

Under perspective projection, a set of parallel lines are projected onto a set of parallel lines in the image plane that meet in a common point. This point of intersection is called the vanishing point (VP). A straight line L in R^3 through the point (b_x, b_y, b_z) can be expressed as $\vec{L} = \vec{a}t + \vec{b}$, where $\vec{a} = (l, m, n)$ is the unit vector giving the direction of L . After the camera project, we get a line on the image plane,

$$x = f \frac{lt + b_x}{nt + b_z}, \quad y = f \frac{mt + b_y}{nt + b_z}, \quad (1)$$

where f is the camera focal length. The VP of the straight line L is the point $P_a = (x_a, y_a)$ on the image plane, where

$$x_a = \lim_{t \rightarrow \infty} x = f \frac{l}{n}, \quad y_a = \lim_{t \rightarrow \infty} y = f \frac{m}{n}. \quad (2)$$

From Eq. (2), we know that 3D parallel lines will intersect at a common vanishing point on the image plane. Because of the limited image size, VP can be located within or outside the image.

The majority of vanishing point detection methods rely on line segments detected in the image, and then the VPs can be computed directly from line segments using cross product operations^[5]. After the lines clustering procedure described in the previous section, we can get three groups of lines that belong to three major mutual orthogonal directions of 3D space. Given n parallel straight lines in one group, the vanishing point coordinates value can be calculated as the average of the $n(n+1)/2$ intersections of the different straight lines on the image plane.

$$\bar{X} = (x_\infty, y_\infty) = \frac{2}{n(n+1)} \sum_{k=1}^{n(n+1)/2} X_k. \quad (3)$$

For a pinhole camera model, the coordinates of a 3D point $\tilde{X} = [X \ Y \ Z]^T$ and its perspective projection image point $\tilde{x}(x, y)$ are represented by

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \\ = P \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = C[R|T] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \quad (4)$$

where P is a 3×4 camera projection matrix of rank 3, R and T are the 3×3 rotation matrix and 3×1 translation vector describing position of camera with respect to world coordinate, C is the camera calibration matrix, it fixes the relation between image pixel coordinates and 3D camera coordinates, the camera intrinsic parameter matrix is represented as

$$C = \begin{bmatrix} d_x & s & u_0 \\ 0 & d_y & v_0 \\ 0 & 0 & 1 \end{bmatrix}, \quad (5)$$

where (u_0, v_0) is the principal point coordinates in image coordinate system, given by the intersection of the optical axis with the image plane. d_x and d_y are the focal lengths in horizontal and vertical pixels, when $d_x/d_y \approx 1$, the camera aspect ratio is close to 1, then the camera skew factor $s \approx 0$.

Let $\vec{v}_i = (l_i, m_i, n_i)_{i=1,2,3}$ be three mutually orthogonal unit vectors, and three VPs on the image plane are $((x_i, y_i)_{i=1,2,3})$. Because $\vec{v}_i (i=1,2,3)$ are mutually orthogonal, we have $\vec{v}_i \vec{v}_j = 0$ ($i \neq j$). Also it can be written as

$$l_i l_j + m_i m_j + n_i n_j = 0, \quad (i \neq j, \quad i, j = 1, 2, 3). \quad (6)$$

After the camera sampling and scanning, we can rearrange Eq. (6) as

$$(u_i - u_0)(u_j - u_0)d_x^2 + (v_i - v_0)(v_j - v_0)d_y^2 + f^2 = 0, \quad (i \neq j, \quad i, j = 1, 2, 3). \quad (7)$$

Define $k = d_y/d_x$, $f_x = f/d_x$, $f_y = f/d_y$, and then Eq. (7) can be rearranged as

$$(u_i - u_0)(u_j - u_0) + (v_i - v_0)(v_j - v_0)k^2 + f_x = 0, \quad (i \neq j, \quad i, j = 1, 2, 3). \quad (8)$$

From Eq. (8), we can get only three equations, so the camera intrinsic parameters cannot be calibrated just by one image with three VPs. If we take another image in a different place, we obtain three more equations similar to Eq. (8). Thus camera intrinsic parameters are calibrated linearly using only three pairs of VPs in two images.

Because the image coordinates of VPs are not affected by the camera pure translation, but are affected only by the camera rotation. The VP v_1 has a direction \vec{V}_1 in the first camera's Euclidean frame, and the VP v_2 has a direction \vec{V}_2 in the second camera's Euclidean frame. We have

$$\begin{aligned} \vec{V}_1 &= C^{-1}v_1 / \|C^{-1}v_1\|, \\ \vec{V}_2 &= C^{-1}v_2 / \|C^{-1}v_2\|. \end{aligned} \quad (9)$$

where C is the camera intrinsic parameter matrix, after three pairs of VPs associated with three mutually orthogonal directions in two views have been calculated, the camera rotation matrix R between two views can be solved by

$$(\vec{v}_2^1, \dots, \vec{v}_2^3) = R(\vec{v}_1^1, \dots, \vec{v}_1^3). \quad (10)$$

Define the two camera projection matrices P and P' as

$$\begin{aligned} P &= C[I|0], \\ P' &= C'[R|T]. \end{aligned} \quad (11)$$

If the same camera is used for two views, then $C = C'$. The correspondence between a set of 3D points

\tilde{X}_i and their image projection points $\tilde{x}_i = [x_i, y_i, 1]$, $\tilde{x}'_i = [x'_i, y'_i, 1]$ between two views is

$$w\tilde{x}_i \cong P\tilde{X}_i, \quad w'\tilde{x}'_i \cong P'\tilde{X}_i. \quad (12)$$

During the VPs detection procedure, we have got three groups of lines that are mutually orthogonal in 3D space. Let OA, OB, OC be three mutually orthogonal lines that intersect at the corner point O to form the Y-shaped corner. And after the camera projection, we can also get a Y-shaped corner on image plane between two views. Let 3D point O with space coordinates $O(X_O, Y_O, Z_O)$, and $o(x_o, y_o)$, $o'(x'_o, y'_o)$ be the image coordinates of the projection of O between two images. The camera translation vector T can be expressed as $T = [T_x, T_y, T_z]$. After the camera intrinsic parameters and the camera rotation matrix have been calculated, we can get three equations about translation vector T , and it can be calculated using least squares estimation method.

After all the camera parameters have been calculated, we can get the camera projection matrices P and P' of the same camera in two views, and it can be used for 3D reconstruction from two images of different views using Eq. (12).

The method of 3D reconstruction using three spatial orthogonal constraints described in this paper has been applied to the real images as shown in Fig. 1; it is an architectural building with two mutual orthogonal planes, and the 3D model is stored in standard virtual reality modelling language (VRML) format.

Three pairs of VPs are statistical calculated, and the detected VPs are shown in Fig. 2.

Having located the VPs positions, we can recover the camera intrinsic parameters by solving Eq. (8) using the least square extrapolation (LSE) method. The calibrated camera parameters are shown in Table 1.

Because $d_x/d_y \approx 1$, the camera aspect ratio is close to 1, then the camera skew factor $s \approx 0$.

The camera projection matrices in two views can be calculated as

$$\begin{aligned} P &= C[I|0] = \begin{bmatrix} 1046.7 & 0 & -54.96 & 0 \\ 0 & 1098.9 & -10.85 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \\ P' &= C[R|T] = \begin{bmatrix} 1036.1 & 32.204 & 155.14 & 363.24 \\ -34.725 & 1098.3 & 10.704 & 231.12 \\ -0.0955 & -0.0135 & 0.9953 & 0.9309 \end{bmatrix}. \end{aligned}$$

Figure 3 is another 3D reconstruction example. This model has five textured planes, and the angle of two 3D reconstructed planes is 91.6° , it is very close to the orthogonal angle.



Fig. 1. (a) Original image a, (b) original image b, (c) 3D reconstruction model.

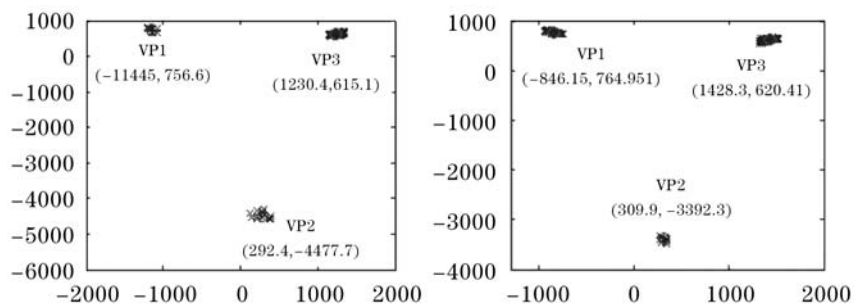


Fig. 2. VPs locations of Figs. 1(a) and (b).



Fig. 3. Original images and the 3D reconstruction model.

Table 1. Camera Intrinsic Parameter

Camera Parameter	f	d_x	d_y	u_0	v_0
(Pixel)	1042.8	1046.7	1098.9	-54.96	-10.85

The simple but powerful constraints of parallelism and orthogonal lines in architectural scenes can be used in 3D reconstruction, but if we cannot find the mutual orthogonal surfaces and as much as possible the parallel lines from the man-made structure surface, the vanishing information cannot be extracted from the images, then we cannot recover the 3D information of the structure.

Y. Shu's e-mail address is shuyuan@mailst.xjtu.edu.cn.

References

1. P. Eisert, E. Steinbach, and B. Girod, IEEE Trans. Circuits and Systems for Video Technology **10**, 261 (2000).
2. A. Marugame, J. Katto, and M. Ohta, IEEE Trans. PAMI **21**, 628 (1999).
3. Z. Y. Zhang, IEEE Trans. PAMI **22**, 1330 (2000).
4. J. Y. Zeng and X. Y. Su, Acta Opt. Sin. (in Chinese) **24**, 628 (2004).
5. J. A. Shufelt, IEEE Trans. PAMI **21**, 282 (1999).