Calibration and 3D Structure Recovery under Varying Cameras using Known Angles

Jong Eun HA*

*Department of Mechanical Engineering
Korea Advanced Institute of Science and Technology
373-1 Kusong-dong, Yusong-gu, Taejon, Korea
jeha@kaist.ac.kr

In So KWEON**

**Department of Electrical Engineering
Korea Advanced Institute of Science and Technology
373-1 Kusong-dong, Yusong-gu, Taejon, Korea
iskweon@kaist.ac.kr

Abstract

In this paper, we present a new algorithm for the calibration of a camera and the recovery of 3D scene structure up to a scale from image sequences using known angles between lines in the scene. The proposed method computes the intrinsic parameters of the camera using the invariance of angles under the similarity transformation. Specifically, we recover the matrix that is the homography between the projective structure and the Euclidean structure using angles. Since this matrix is a unique one in the given set of image sequences, we can easily deal with the problem of varying intrinsic parameters of the camera.

Experimental results on the synthetic and real images demonstrate the feasibility of the proposed algorithm.

1 Introduction

The recovery of 3D scene structure from image sequences has been an active research topic in computer vision. A classical stereo approach first calibrates the camera and then recovers 3D scene structure using the correspondences in two images. Most calibration methods [1, 2] compute the relative positions of the cameras and their intrinsic parameters using 3D coordinates of points on a known calibration target. However, it is almost impossible to use the same calibration target for the wide range of vision tasks that require cameras with long focal lengths for magnification as well as short ones for a larger field-of-view. Moreover, many robotic applications demand cameras to be calibrated on-line, which makes impossible to set out specific calibration targets for different camera set-ups.

Faugeras [3] presented a calibration method using the known angle between lines in the scene. First, he derived the constraint for the image of the absolute conic using the known angle between two lines that pass through the optical center and it is a quadratic on the unknown coefficients of the image of the absolute conic. To extend this result to the known angle between lines in 3D, it is required to recover the given scene up to affine transformations.

Faugeras *et al.* [4] also presented a stratified approach to recover the Euclidean geometry of a scene by using *a priori* information about the scene. They first used *a priori* information about the affine structure of the scene such as 3-D parallel lines and ratios of lengths of parallel 3-D segments. To recover the metric structure of the scene they used *a priori* information such as 3-D angles and ratios of lengths of 3-D segments. They recovered the scene structure sequentially from the projective to the affine and finally to the Euclidean.

Our proposed algorithm recovers directly the homography matrix between the projective and the Euclidean structure, without going through an intermediate step of affine reconstruction.

Boufama *et al.* [5] presented a method for directly estimating the transformation matrix between the projective solution and the Euclidean one using constraints in the scene. They used fixating a point, laying on the horizontal plane, vertical alignment and the distant between points in 3-D.

In this paper, we present a calibration method that uses a constraint – angle between two lines – in the scene rather than a calibration target. The angle invariance under the similarity transformation is used to derive the constraint for the homography matrix between the projective structure and the Euclidean structure. This homography is a unique one in the given sets of images. By this homography we can deal with the varying intrinsic parameters of camera only using the projective reconstruction.

2 Reconstruction by Homography Matrix

The process of projection of a point in 3-D to the image plane can be expressed as a linear matrix operation in the homogenous coordinates. First, there is a rigid body transformation between the world coordinates $\mathbf{X}_w$ and the camera-centered coordinates $\mathbf{X}_c$. The next stage is perspective projection of $\mathbf{X}_c$ onto $\mathbf{x}$ in the image plane. Finally, the image coordinates $\mathbf{x}$ are converted to the pixel coordinates $\mathbf{m} = (u, v, 1)^T$. These processes can be represented as:

$$\mathbf{m} = \mathbf{P}_{euc}\mathbf{X}_w \tag{1}$$

where

$$\mathbf{P}_{euc} = \mathbf{A}\mathbf{P}_0\mathbf{T}$$
$$= \begin{bmatrix} a_u & g & u_0 \\ 0 & a_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}_3^T & 1 \end{bmatrix}$$

$\mathbf{P}_{euc}$ is the camera projection matrix for a perspective camera. a_u and a_v are the scale aspect of the x and y axis in the image, g is a skew factor, and u_0 and v_0 are the pixel coordinates of the principal point. We assume that the skew in the camera is negligible:

$$g = 0 \tag{2}$$

It is well known that projective reconstruction can be obtained from uncalibrated image sequences [8, 9]. In this case the projection equation is given by

$$\mathbf{m}_j^i \cong \mathbf{P}_{proj}^i \mathbf{Q}\mathbf{Q}^{-1}\mathbf{X}_{proj}^j \tag{3}$$

where $\mathbf{m}_j^i$ is a 3x1 vector representing the j-th point in the i-th image, $\mathbf{P}_{proj}^i$ is a 3x4 projective projection matrix of i-th camera, $\mathbf{Q}$ is a 4x4 nonsingular homography matrix in P^3 , and $\mathbf{X}_{proj}^j$ is a 4x1 projective structure in P^3 .

In Eq. (3), there exist many nonsingular matrices $\mathbf{Q}$ which satisfy the equation. Among many $\mathbf{Q}$, there exist a unique $\mathbf{Q}$ matrix that transforms the projective structure to the Euclidean structure. Through this unique $\mathbf{Q}$ matrix we can calibrate each camera and reconstruct a 3-D scene up to the similarity transformation. Such a $\mathbf{Q}$ matrix satisfies the following relations:

$$\begin{aligned} \mathbf{P}_{euc}^i &\cong \mathbf{P}_{proj}^i \mathbf{Q} \\ \mathbf{X}_{euc}^j &\cong \mathbf{Q}^{-1}\mathbf{X}_{proj}^j \end{aligned} \tag{4}$$

where $\mathbf{X}_{euc}^j$ is a metric structure of a 3-D point.

If we set the world coordinate system at the optical center of the first camera, we can obtain the projective projection matrix and the Euclidean projection matrix as $\mathbf{P}_{proj}^0 = [\mathbf{I}_{3X3} \ \ \mathbf{0}_3]$ and $\mathbf{P}_{euc}^0 = [\mathbf{A}_0 \ \ \mathbf{0}_3]$. If we substitute these relations into Eq. (4), we can obtain the elements of $\mathbf{Q}$ matrix:

$$\begin{aligned} &\mathbf{P}_{euc}^0 \cong \mathbf{P}_{proj}^0 \mathbf{Q} \Leftrightarrow [\mathbf{A}_0 \ \ \mathbf{0}_3] \cong [\mathbf{I}_{3X3} \ \ \mathbf{0}_3]\mathbf{Q} \\ &\Leftrightarrow \exists (q_1 \ q_2 \ q_3 \ q_4) \ \Big| \ \mathbf{Q} \cong \begin{bmatrix} \mathbf{A}_0 & \mathbf{0}_3 \\ q_1 \ q_2 \ q_3 & q_4 \end{bmatrix} \end{aligned} \tag{5}$$

Since $\mathbf{Q}$ is defined up to a scale, we can represent it as:

$$\mathbf{Q} = \begin{bmatrix} a_u & 0 & u_0 & 0 \\ 0 & a_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \\ q_1 & q_2 & q_3 & 1 \end{bmatrix} \tag{6}$$

After computing $\mathbf{Q}$, we can obtain an Euclidean structure up to a global scale and can compute the rotation and translation between cameras through the decomposition of the Euclidean projection matrix computed using Eq. (4).

3 Camera Calibration

In this section, we describe a calibration algorithm using *a priori* known angles between two lines in 3-D. Let us assume three points A, B and C are specified with their world coordinates $\mathbf{X}_w^A, \mathbf{X}_w^B$ and $\mathbf{X}_w^C$ as shown in Fig. 1. It is a well-known result that from image correspondences, $\{\mathbf{m}_i^1\}$ and $\{\mathbf{m}_i^2\}$, we can reconstruct a given scene up to the projective transformation [6, 7]. If we set the world coordinate at the optical center of the first camera and if we know $\mathbf{Q}$ matrix in Eq. (3), we can compute the Euclidean structure, $s\mathbf{X}_w^A$, $s\mathbf{X}_w^B$ and $s\mathbf{X}_w^C$, up to a global scale.

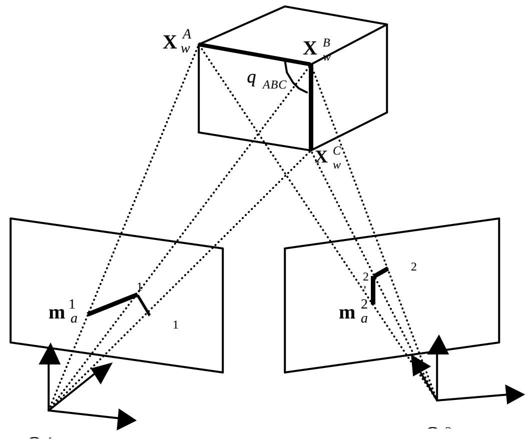


Fig. 1 The configuration of the imaging system

Since the Euclidean structure is recovered up to a scale, angle formed by the three points is invariant under the similarity transformations:

$$\cos\theta_{ABC} = \frac{\vec{\mathbf{BA}} \bullet \vec{\mathbf{BC}}}{\left|\vec{\mathbf{BA}}\right|\left|\vec{\mathbf{BC}}\right|} = \frac{\vec{\mathbf{B'A'}} \bullet \vec{\mathbf{B'C'}}}{\left|\vec{\mathbf{B'A'}}\right|\left|\vec{\mathbf{B'C'}}\right|} = \cos\theta_{A'B'C'} \qquad (7)$$

where $\vec{\mathbf{BA}} = \mathbf{X}_w^A - \mathbf{X}_w^B$, $\vec{\mathbf{BC}} = \mathbf{X}_w^C - \mathbf{X}_w^B$, $\vec{\mathbf{B'A'}} = s\mathbf{X}_w^A - s\mathbf{X}_w^B$ and $\vec{\mathbf{B'C'}} = s\mathbf{X}_w^C - s\mathbf{X}_w^B$.

Set $\mathbf{Q}^{-1}$ and the projective structure of j-th point as:

$$\mathbf{Q}^{-1} = \begin{pmatrix} \alpha_u & 0 & u_0 & 0 \\ 0 & \alpha_v & v_0 & 0 \\ 0 & 0 & 0 & 0 \\ q_1 & q_2 & q_3 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{Q}_1^{-T} \\ \mathbf{Q}_2^{-T} \\ \mathbf{Q}_3^{-T} \\ \mathbf{Q}_4^{-T} \end{pmatrix} \quad \mathbf{X}_{proj}^j = \begin{pmatrix} p_j \\ q_j \\ r_j \\ s_j \end{pmatrix} (j = A, B, C) \quad (8)$$

where $\mathbf{Q}_i^{-T}$ is the i-th row of $\mathbf{Q}^{-1}$.

By substituting Eq. (8) into Eq. (3), the following equation can be derived:

$$\mathbf{X}_{euc}^j = \begin{pmatrix} X_{euc}^j \\ Y_{euc}^j \\ Z_{euc}^j \end{pmatrix} = \begin{pmatrix} \dfrac{\mathbf{Q}_1^{-T} \bullet \mathbf{X}_{proj}^j}{\mathbf{Q}_4^{-T} \bullet \mathbf{X}_{proj}^j} \\ \dfrac{\mathbf{Q}_2^{-T} \bullet \mathbf{X}_{proj}^j}{\mathbf{Q}_4^{-T} \bullet \mathbf{X}_{proj}^j} \\ \dfrac{\mathbf{Q}_3^{-T} \bullet \mathbf{X}_{proj}^j}{\mathbf{Q}_4^{-T} \bullet \mathbf{X}_{proj}^j} \end{pmatrix} \quad (j = A, B, C) \qquad (9)$$

Using Eq. (9), we can represent $\cos\theta_{ABC}$ as a function of the parameters of **Q**:

$$\cos\theta_{ABC} = \frac{\vec{\mathbf{BA}} \bullet \vec{\mathbf{BC}}}{\left|\vec{\mathbf{BA}}\right|\left|\vec{\mathbf{BC}}\right|} = f(\alpha_u, \alpha_v, u_0, v_0, q_1, q_2, q_3) \qquad (10)$$

Therefore, the scene constraint – the angle invariance under the similarity transformation – can be translated to a constraint on the intrinsic parameters. Specifically, from Eq. (10), we compute the Euclidean projection matrix from the projective one. A nonlinear least squares method computes an approximate solution by minimizing the following criterion:

$$\min \sum_{i=1}^{N} \left(\cos(\theta_{ABC}) - \frac{\vec{\mathbf{B}_i\mathbf{A}_i} \bullet \vec{\mathbf{B}_i\mathbf{C}_i}}{\left|\vec{\mathbf{B}_i\mathbf{A}_i}\right|\left|\vec{\mathbf{B}_i\mathbf{C}_i}\right|} \right)^2 \qquad (11)$$
$$= \min \sum_{i=1}^{N} \left(\cos(\theta_{ABC}) - f_i(\alpha_u, \alpha_v, u_0, v_0, q_1, q_2, q_3)\right)^2$$

Here, we use the Levenberg-Marquardt method [8] to solve the nonlinear minimization problem. The initial estimate of unknowns is computed using the method of Bougnoux [9]. The scaled Euclidean structure follows easily.

The algorithm for camera calibration using the angle invariance is summarized in Table 1.

Table 1 Calibration algorithm using known angles

1. Find the projections $\mathbf{P}_{proj}^i$ by projective reconstruction between images, e.g., 1-2, . .,.1-N.
2. Obtain an initial estimate of unknowns $(\alpha_u, \alpha_v, q_1, q_2, q_3)$ using the projective reconstruction between images 1 and 2.
3. Select N sets of three control points in the scene.
4. Find a 4x4 homography **Q** by solving Eq. (11).
5. Find the Euclidean projections by $\mathbf{P}_{euc}^i \cong \mathbf{P}_{proj}^i \mathbf{Q}$.
6. Recover the Euclidean structure by $\mathbf{X}_{euc}^j \cong \mathbf{Q}^{-1}\mathbf{X}_{proj}^j$.

4 Experimental Results

We have tested the proposed algorithm using the synthetic and real images. We used a synthetic image sequence of a 15x15x15cm^3 hexahedron with 9 control points on each three planes as shown in Fig. 2. Images were taken from six different viewing positions as described in Table 3. The influence of image noise was evaluated by adding a white Gaussian noise, with the standard deviation from 0.2 to 1.5 pixels, to the control points. The internal parameters of the synthetic camera were varied as shown in Table 2. Fig. 3 shows the mean of the relative error of the internal parameters over 50 trial runs for the second and the fifth camera. Fig. 3-(a) and 3-(b) correspond to the second and fifth camera's estimated intrinsic parameter and the value of the intrinsic parameter varied according to Table 2. Proposed algorithm effectively cope with varying cameras. As expected, the error in the estimates of internal parameters increases monotonically as the image noise increases.

The true value of $(q_1, q_2, q_3)^3$ is obtained by computing the homography of Eq. (4) between the known projective and Euclidean structures.

The mean of the relative error of the external parameters is depicted in Fig. 4. The estimated Euclidean projection matrix is decompos ed into the rotational and transnational components. As in the internal parameters, a similar linear tendency to image noise can be observed.

Fig. 5 depicts the accuracy of the method in terms of 3-D structure recovery. From the scaled Euclidean structure, we estimated the relative error of the angle formed by three adjacent control points and the ratio of length between two control points. The accuracy in the estimates of the angle degrades monotonically as image noise increases. However, it is interesting to note that the accuracy in the ratio of length is preserved nicely regardless of noise level.

Table 2 The intrinsic parameters used for the varying camera.

	Camera 1	Camera 2	Camera 3	Camera 4	Camera 5
α_u	700.0	750.0	650.0	800.0	600.0
α_v	700.0	750.0	650.0	800.0	600.0
u_0	256.0	256.0	256.0	256.0	256.0
v_0	256.0	256.0	256.0	256.0	256.0

Table 3 The extrinsic parameters used in the generation of the synthetic image sequence

	Rotation $(\theta_x, \theta_y, \theta_z)$ [deg]	Translation (t_x, t_y, t_z) [cm]
Image 1 - 2	$(7^\circ, -5^\circ, 10^\circ)$	$(2.5,\ 5.0,\ 2.5)$
Image 1- 3	$(-2^\circ, -10^\circ, -15^\circ)$	$(5.0,\ -5.0, 5.0)$
Image 1- 4	$(3^\circ, -15^\circ, 20^\circ)$	$(7.5,\ 5.0,\ 7.5)$
Image 1- 5	$(-5^\circ, -20^\circ, -10^\circ)$	$(9.0,\ -5.0, 9.0)$

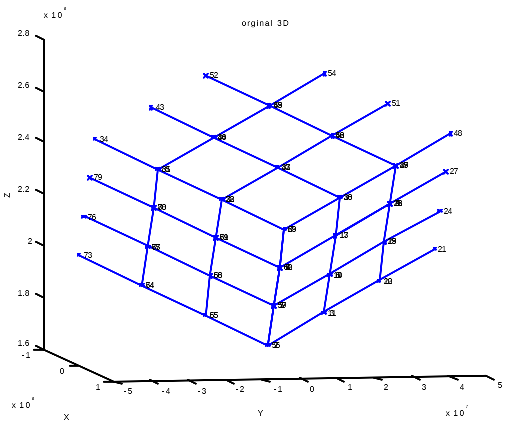
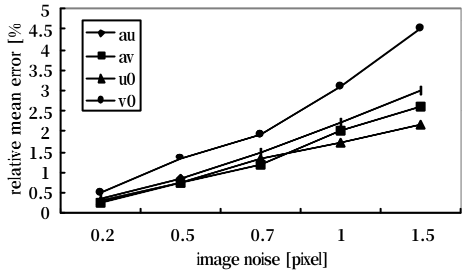
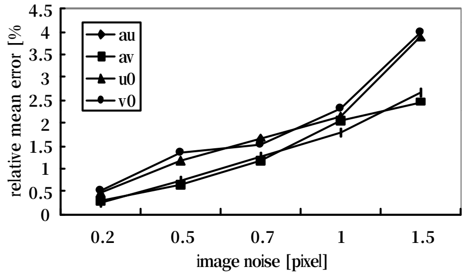


Fig. 2 The 3-D structure of a hexahedron used in the synthetic experiment: the numbered points indicate the control points

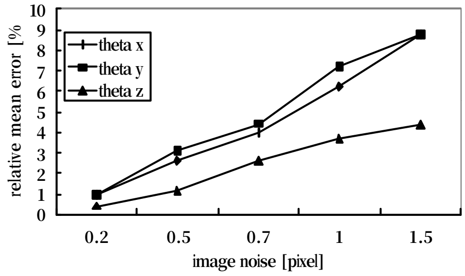


(a)

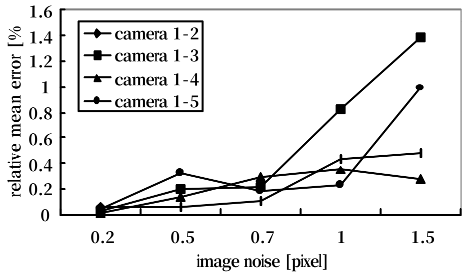


(b)

Fig. 3 The relative mean error of the estimated intrinsic parameters (a) intrinsic parameters of the second camera (b) intrinsic parameters of the fifth camera

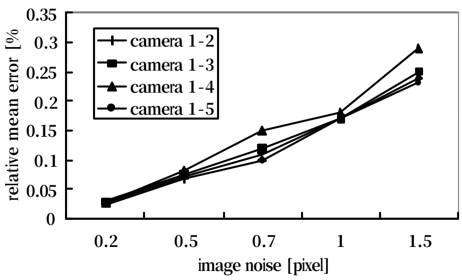


(a)

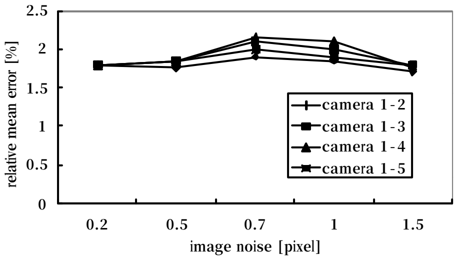


(b)

Fig. 4 The relative mean error of the estimated extrinsic parameter (a) the rotation error between cameras 1 and 2 (b) the error in the translation direction.



(a)

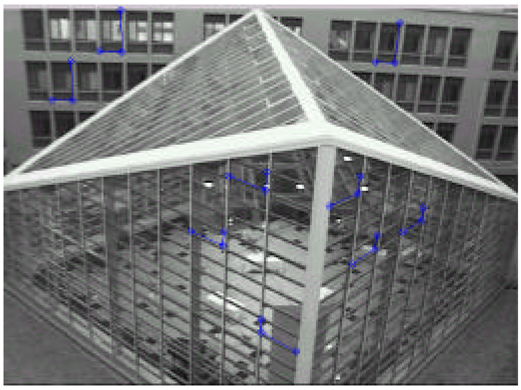


(b)

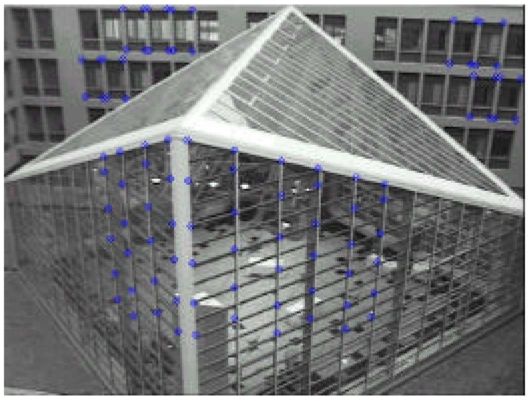
Fig. 5 The relative mean error of the angle between control points and the length ratio (a) the angle between control points (b) the length ratio.

Fig. 6 shows the images of an outdoor building scene used for verifying the effectiveness of our method. The image sequence is obtained by using a hand-held camcoder. The corresponding points between images are manually selected. The control angles between two lines are assumed to be 90°. We used 9 control angles, which is shown in Fig. 6-(a), to compute the Q matrix. Fig. 6-(b) shows all the corresponding points used in the reconstruction.

The mean and standard deviation of angles formed by the control points from the estimated 3-D structure are $92.2^\circ/3.27^\circ$. It is a good estimate since the window frames are manufactured to have the right angle. The rotation angles between image 1 and 2 are $(q_x, q_y, q_z) = (-6.75^\circ, -13.3^\circ, -3.96^\circ)$ and the estimated unit translation direction is $(T_x, T_y, T_z) = (0.984, 0.00976, 0.178)$. It can be seen that these estimates are reasonable according to the images 1 and 2. Fig. 7-(a) shows the estimated 3-D structure of all the corresponding ponts. The proposed algorithm recovers an Euclidean structure up to a global scale, thus we arbitrarly select this scale for display. Fig. 7-(b) shows a bird's-eye view of the estimated 3-D structure. We can observe that the coplanarity of the original structure is well preserved in the recovered structure.

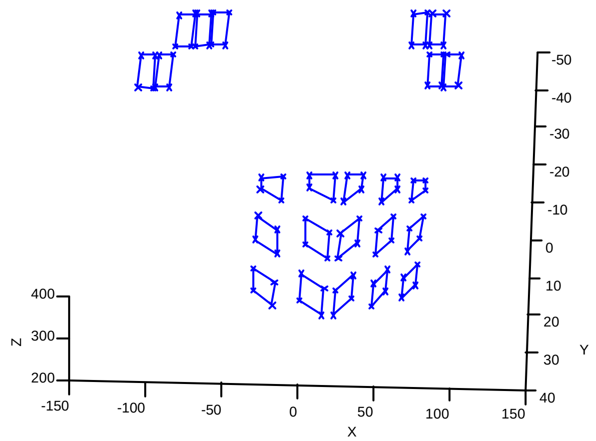


(a)

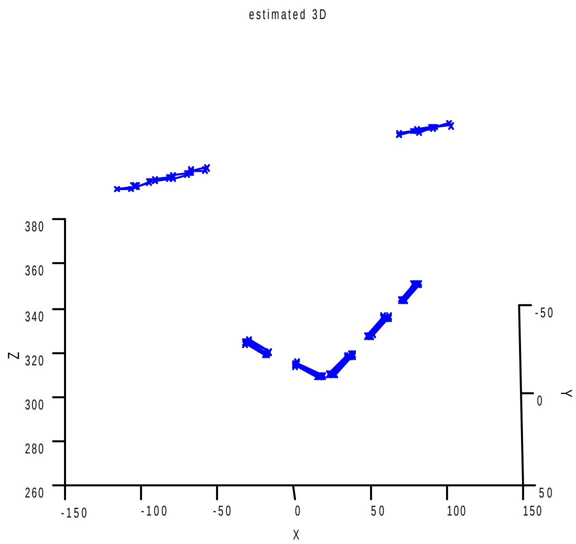


(b)

Fig. 6 The building image sequences: (a) image 1 and control angles formed by three neighboring points (b) image 2 and all the corresponding points used in the reconstruction.



(a)



(b)

Fig. 7 The recovered 3D structure by proposed algorithm (a) estimated 3-D structure of all the corresponding points (b) the bird's-eye view of (a).

5 Conclusion

We have presented a calibration algorithm using a specific scene constraint – the invariance of angles under similarity transformations – while recovering the Euclidean structure up to a scale from image correspondences. It is well known that a non-singular transformation matrix **Q**, whose elements consist of the intrinsic parameters of the camera, translates the projective reconstruction to the Euclidean one. Based on this property of **Q** matrix, our method exploits the invariance characteristic of angles under the similarity transformation for the camera calibration and the 3-D structure recovery. Since the strong scene constraint is applied to the minimization process to compute the camera parameters, the estimated Euclidean structure automatically satisfies the scene constraint. Furthermore, it can effectively cope with the varying intrinsic parameters of the camera by only using projective reconstruction.

One limitation is, however, that the method needs *a priori* 3-D scene information. We envision that this method will be effective to the scene consisting of man-made objects, such as buildings.

References

[1] R. Y. Tsai, "A Versatile Camera Calibration Technique for High Accuracy 3D Machine Vision Metrology using Off-the-shelf TV Cameras and Lenses," *IEEE Journal of Robotics and Automation*, Vol. 3, No. 4, Aug, 1987

[2] O.D. Faugeras and G. Toscani, "Camera Calibration for 3D Computer Vision," *Proceedings of International Workshop on Machine Vision and Machine Intelligence, Tokyo, Japan*, 1987

[3] O. Faugeras, "Stratification of Three-dimensional Vision: Projective, Affine and Metric representations," *J. Opt. Soc. Am. A*, Vol.12, No. 3, pp 465-484, 1995.

[4] O. Faugeras, S. Laveau, L. Robert, G. Csurka and C. Zeller, "3-D Reconstruction of Urban Scenes from Sequences of Images," *INRIA RR-2572*, 1995.

[5] B. Boufama, R. Mohr and F. Veillon, "Euclidean Constraints for Uncalibrated Reconstruction," *International Conference on Computer Vision*, pp. 466-470, 1993.

[6] O. Faugeras, "What Can Be Seen in Three Dimensions with an Uncalibrated Stereo Rig?," *Proceedings of the 2nd European Conference on Computer Vision*, pp. 563-578, 1992.

[7] R. Hartley, R. Gupta, and T. Chang, "Stereo from Uncalibrated Cameras," *Proceedings of IEEE Int. Conference on Computer Vision and Pattern Recognition*, pp. 761-764, 1992.

[8] W. H. Press, B. P. Flannery, S. A. Teukolsky and W. T. Vetterling, *Numerical Recipes in C*, Cambridge University Press.

[9] S. Bougnoux, "From Projective to Euclidean Space under Any Practical Situation, a Criticism of Self-calibration," *Proceedings of International Conference on Computer Vision*, pp. 790-796, 1998.