A Linear Shape from Motion Algorithm using Rotation Information of the Cameras

Akira Amano, Graduate School of Informatics, Kyoto University
Tsuyoshi Migita, Faculty of Information Science, Hiroshima City University
Naoki Asada, Faculty of Information Science, Hiroshima City University

Abstract Although Shape from Motion problem is formulated as nonlinear least squares problem, it is generally difficult to solve without restrictions on the scene or the motion. If the rotation information of the cameras are known, estimation of the scene structure and the camera translation is linearized. The effectiveness of the method is shown with the real image sequence with the angle information.

Keywords: Shape from Motion, rotation information, linearization, eigen decomposition

1 Introduction

One of the main problem of Computer Vision is to recover 3D shape information from 2D image information. The problem of recovering the 3D shape and the camera positions from the multiple images is called Shape from Motion (SfM) problem. The input of the problem is 2D coordinates of each feature points in every image, and the output is their 3D positions and camera positions of each image. As the 2D coordinates of each feature point is a projection of the 3D position into each image plane, the problem becomes inverse problem of the projection, and as the 2D projection of 3D feature point is formulated as a nonlinear equation, the problem is formulated as a nonlinear optimization[1].

Difficulty of the problem vastly changes according to the property of the input image set. In the previous works, there are implicit or explicit restrictions to the image set such as, distance between the object and the camera is relatively large[2], entire object is captured in the every image[2, 1], or the camera translation or rotation is relatively small[3].

However when we think of recovering a large object such as a building, we can not always satisfy these restrictions. Especially when we think of recovering the entire shape of them, satisfying these restrictions become very difficult in urban situations. One way of coping with this problem is to incorporate non-image information.

In this paper, we incorporate the angle information of each camera for recovering 3D shape of a large object from the multiple images. These angle information is obtained with a angle sensor or a tripod with certain angle measure. With the angle information, we can recover 3D shape by linear calculation without any approximation. This method has advantage that it has no restrictions to the image properties.

2 Shape from Motion

Shape and motion problem is described by the following variables. Note that the camera focal length l is known and constant.

- Shape $\boldsymbol{s}_p = (x_p, y_p, z_p)^T, 1 \leq p \leq P$

  We have P feature points in the scene, and the 3D coordinates of the pth feature point on the object coordinate system XYZ is described as above.

- Rotation matrix $R_f, 1 \leq f \leq F$

  Rotation matrix R_f expresses the fth frame rotation of the object coordinate system viewed from the camera coordinate system.

- Translation vector $\boldsymbol{t}_f = (t_{fx}, t_{fy}, t_{fz}), 1 \leq f \leq F$

  Translation vector $\boldsymbol{t}_f$ expresses the fth frame translation of the object coordinate system viewed from the camera coordinate system.

With above notations, 3D coordinates of the pth feature point viewed from the camera coordinate system of the fth frame become as follows.

$$\boldsymbol{s}_{fp} = R_f \boldsymbol{s}_p + \boldsymbol{t}_f.$$

When the pth feature point at the fth frame $\boldsymbol{s}_{fp}$ is projected to the image, 2D coordinate $\boldsymbol{u}_{fp} = (u_{fp}, v_{fp})^T$ becomes as follows,

$$\boldsymbol{u}_{fp} \quad = \quad \mathcal{P}(\boldsymbol{s}_{fp}) = \mathcal{P}(R_f \boldsymbol{s}_p + \boldsymbol{t}_f) \qquad (1)$$

where $\mathcal{P}$ represents the perspective projection operator defined as follows.

$$\mathcal{P}\left[\begin{array}{c} x \\ y \\ z \end{array}\right] := \frac{l}{z}\left[\begin{array}{c} x \\ y \end{array}\right]$$

In the SfM problem, the coordinates of the pth feature point in the fth frame is given as $\tilde{\boldsymbol{u}}_{fp}$. The problem is summarized as finding the object shape $\boldsymbol{s}_p$ and the camera rotation and translation $R_f, \boldsymbol{t}_f$ which best fit the above feature point coordinates $\tilde{\boldsymbol{u}}_{fp}$. In the most simple form, the problem is formulated as the nonlinear optimization of following formula.

$$\arg \min_{R_f, \boldsymbol{t}_f, \boldsymbol{s}_p} \sum_{(f,p)} |\mathcal{P}(R_f \boldsymbol{s}_p + \boldsymbol{t}_f) - \tilde{\boldsymbol{u}}_{fp}|^2 \qquad (2)$$

In the case that every feature points are viewed in every image, this optimization becomes relatively easy problem, however, in the case that only small number of feature points are viewed in each image, it becomes highly difficult, which means that many local minima occurs in the optimization process. In most previous methods of SfM, some sort of restrictions are assumed implicitly or explicitly[4, 3, 5, 6].

3 Incorporating Angle Information

In the case that the only small number of feature points are viewed in each image, we can think of incorporating non-image information to recover shape and motion. Here we think of incorporating angle information.

3.1 Formulation

Incorporating angle information means, the rotation matrices R_f corresponding to the images are known. It is known that by incorporating such information, SfM problem becomes linear[7]. Here, we consider a function of eq.(2) multiplied by the following projective depth λ_{fp}.

$$\lambda_{fp} = R_{f2}^T \boldsymbol{s}_p + t_{fz} \qquad (3)$$

where, R_{f2} corresponds to the third row of the matrix R_f. By incorporating R_f information, the evaluation function becomes linear as shown below.

$$\begin{aligned} & \sum_{(f,p)} \lambda_{fp}^2 |\mathcal{P}(R_f \boldsymbol{s}_p + \boldsymbol{t}_f) - \tilde{\boldsymbol{u}}_{fp}|^2 \\ = & \sum_{(f,p)} \left\| \left[\begin{array}{ccc} 1 & 0 & -\tilde{u}_{fp} \\ 0 & 1 & -\tilde{v}_{fp} \end{array}\right] [R_f | I] \left[\begin{array}{c} \boldsymbol{s}_p \\ \boldsymbol{t}_f \end{array}\right] \right\|^2 \qquad (4) \end{aligned}$$

Let $\boldsymbol{x} = [\boldsymbol{s}_1^T\ \boldsymbol{s}_2^T ... \boldsymbol{s}_P^T\ \boldsymbol{t}_1^T\ \boldsymbol{t}_2^T ... \boldsymbol{t}_F^T]^T$, the problem is written as follows.

$$\arg \min_{\boldsymbol{x}}\ \boldsymbol{x}^T A \boldsymbol{x}$$

Here, matrix A becomes block diagonal matrix as follows.

$$A = \left[\begin{array}{ccc|ccc} S_1 & & & U_{11} & \cdots & U_{1F} \\ & \ddots & & \vdots & & \vdots \\ & & S_P & U_{P1} & \cdots & U_{PF} \\ \hline & & & T_1 & & \\ & \text{Sym.} & & & \ddots & \\ & & & & & T_F \end{array}\right]$$

Each element of A becomes as follows.

$$\begin{aligned} S_p &= \sum_f R_f^T P_{fp} R_f & (5) \\ U_{fp} &= R_f^T P_{fp} & (6) \\ T_f &= \sum_p P_{fp} & (7) \end{aligned}$$

where P_{fp} is as follows.

$$P_{fp} = \left[\begin{array}{ccc} 1 & 0 & -u_{fp} \\ 0 & 1 & -v_{fp} \\ -u_{fp} & -v_{fp} & u_{fp}^2 + v_{fp}^2 \end{array}\right]$$

It is well known that the eigenvector of A which corresponds to the smallest eigenvalue gives optimal solution of x.

4 Calculation method

As we need only one eigenvector corresponding to the smallest eigenvalue as the solution, it is efficient to use the inverse power method to achieve it. This inverse power method gives fastest computation time compared to the other methods such as the Jacobi Transform method. Using the random vector $\boldsymbol{x}_0$ as the initial value, we can obtain the eigenvector by performing the following iteration for 5 or 6 times.

$$\boldsymbol{x}_{k+1} = A^{-1} \boldsymbol{x}_k$$

5 Experimental Results

5.1 Simulation Results

We first employed simulation experiments to show the effectiveness of our method. Three real image sequences are used for the experiment. The size of each image is 1024 × 768 pixels, and the focal length is 1037 in pixels. The coordinates of each feature

Table 1: Characteristics of three image sequences for the simulation experiments. N is the number of images, P is the number of feature points, and r is the average number of images each feature point appears.

	seq.1	seq.2	seq.3
N	198	29	144
P	300	124	239
r	10.0	5.4	11.0
range	mid range	mid range	close range

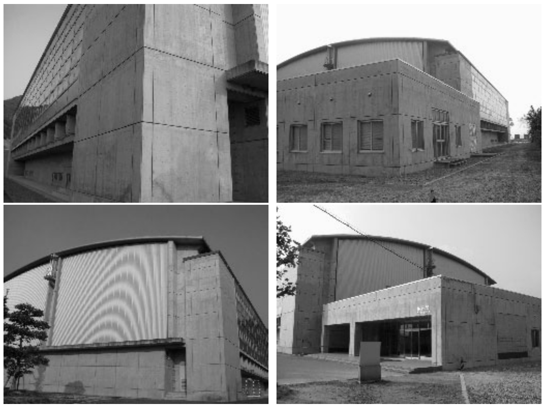

Figure 1: Images of the gym. 4 out of 198 images are shown.

points are tracked manually and the angle information of the cameras are obtained by the manual optimization of eq.(2). In the experiment, the random gaussian error whose average is 0 and standard deviation is σ degree is added to the horizontal and vertical angle and the rotation around the optical axis information. We used 2.0 and 5.0 for the σ which were the typical values for the angle sensors and the mechanical measurement systems.

The charactersitics of each image sequence is shown in Tab.1. The sequence 1 is an large data set with the average frames for each feature point appears is about 10. In the sequence 2, this number decreases to 5.4 which makes the optimization difficult. The sequence 3 has near average frames for each feature point appear, however, the images are taken very close to the building which make the problem very difficult.

5.1.1 Experiment on sequence 1

Fig.1 shows four images out of 198 frames of sequence 1. Number of feature point was 300 and average number of images each feature point appears was about 10. Fig.2 is top view of true shape where rectangular shape in the center corresponds to the shape of the building and surrounding points correspond to

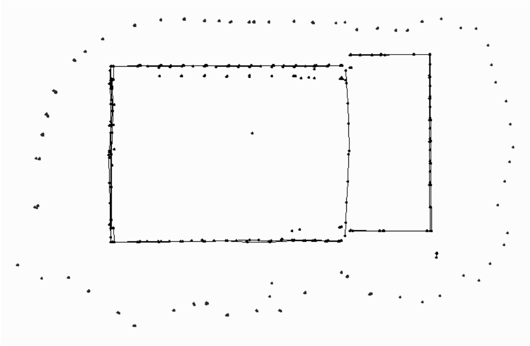

Figure 2: True shape.

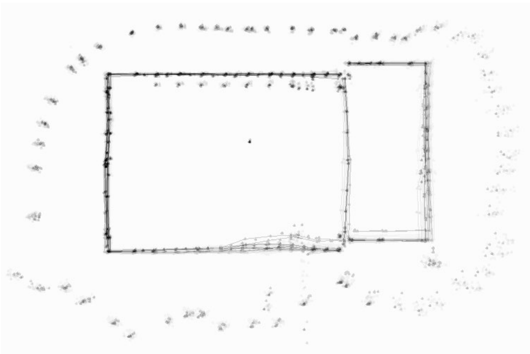

Figure 3: Estimated shape with $\sigma = 2$.

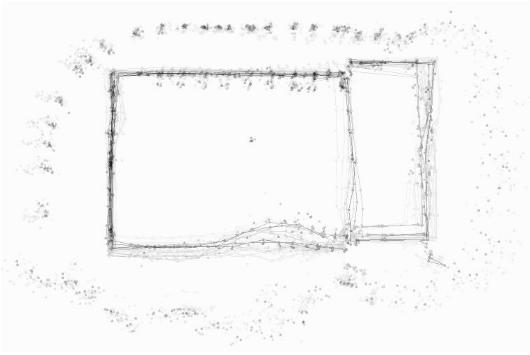

Figure 4: Estimated shape with $\sigma = 5$.

the camera positions.

Fig.3 shows sum of 20 experimental results of the top views of recovered shape with different angle error of standard deviation $\sigma = 2$ degree added. Fig.4 shows result of $\sigma = 5$ error added. Note that the recovering shape has distortions in vertical direction also. In the case of $\sigma = 2$, the recoverd shape has small distortions while it becomes slightly large in the case of $\sigma = 5$. In this case, both results can be used as an input for the bundle adjustment.

5.1.2 Experiment on sequence 2

Fig.5 shows four images out of 29 in sequence 2. The number of feature point is 124 and the average number of images each feature point appears was about 5.4. Fig.6 shows the top view of the true shape.

Fig.7 shows the recovered images with $\sigma = 2$. As there exist feature points which only appears in the two images, these points are recovered with large errors. As the the average number of images each feature point appears is small compared to sequence 1, the SfM problem is difficult compared to sequence 1. However, except for above outliers, recovered shape shows good estimation which shows effectiveness of using angle information together with image information.

Fig.8 shows recovered image with error of $\sigma = 5$. Although the shape is largely distorted, it is still usefull for initial values for bundle adjustment because it has no topological error such as in Fig.9 which were obtained by the direct optimization of equation (2). This kind of topological errors are fatal for the initial values of the bundle adjustment.

5.1.3 Experiment on sequence 3

Fig.10 shows four images out of 144 in sequence 3. The number of feature points is 239 and the average number of images each feature point appears was about 11. Note that, in this sequence, some images are taken very close to the building so the captured area in some images are very small part of entire building.

Fig.11 shows recovered image with $\sigma = 2$. As in the result of previous experiment, this result also shows no topological error. However, the position errors are slightly large. This is because the resulting position error becomes relatively large with close range images. This leads to the point that the far distance images are prefereable for the SfM problem, however, the near distance images are still sufficient for obtaining the initial values for the bundle adjustment.

Fig.12 shows recovered image with $\sigma = 5$. The shape is largely distorted, however, it is still usefull for initial values for the bundle adjustment.

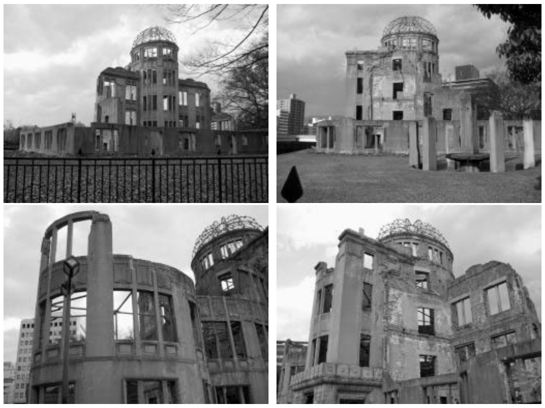

Figure 5: Images of the Atomic Bomb Dome. 4 out of 29 images are shown.

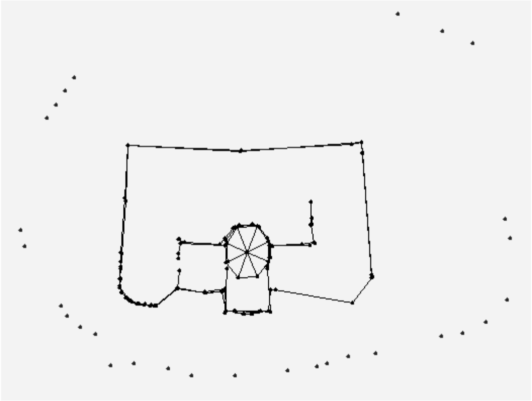

Figure 6: True shape.

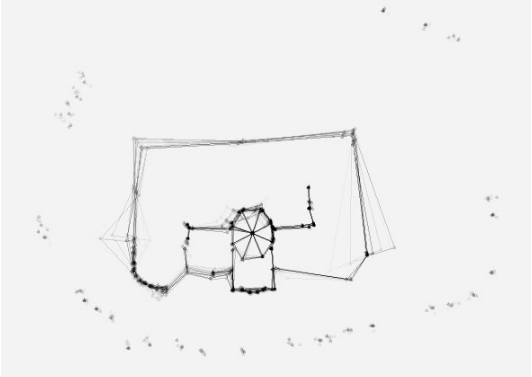

Figure 7: Estimated shape with $\sigma = 2$.

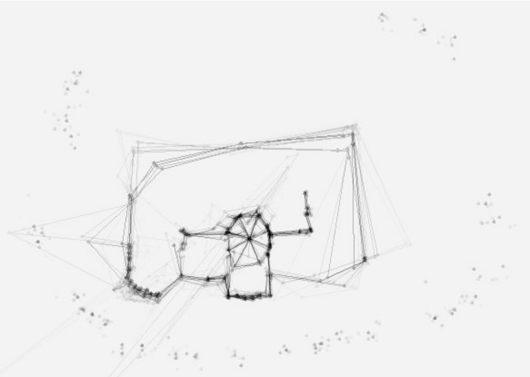

Figure 8: Estimated shape with $\sigma = 5$.

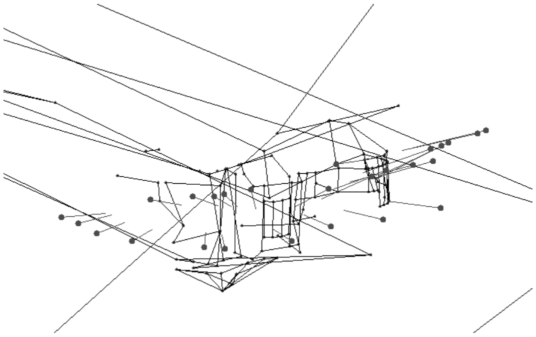

Figure 9: Failed estimation in nonliner method.

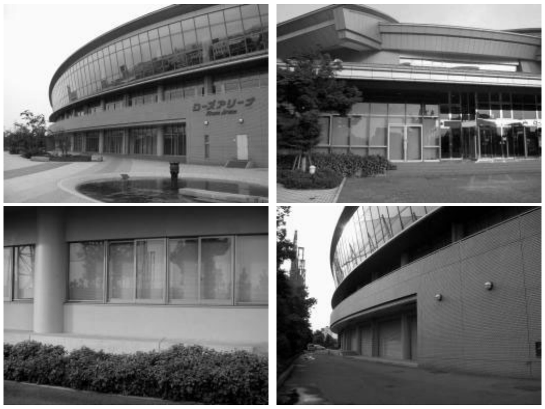

Figure 10: Images of the Stadium. 4 images out of 144 ones are shown.

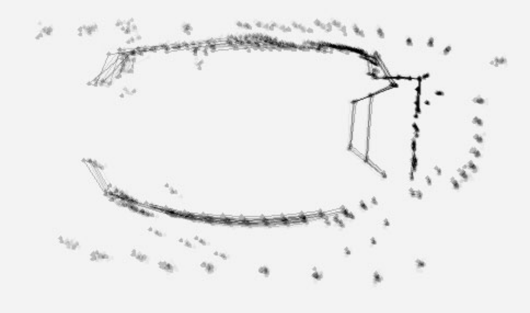

Figure 11: Estimated shape with $\sigma = 2$.

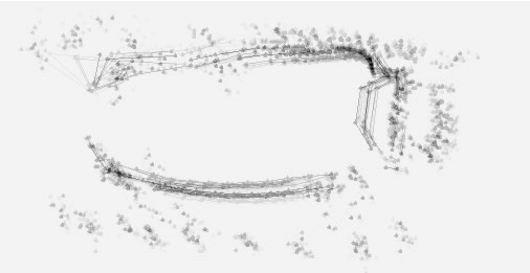

Figure 12: Estimated shape with $\sigma = 5$.

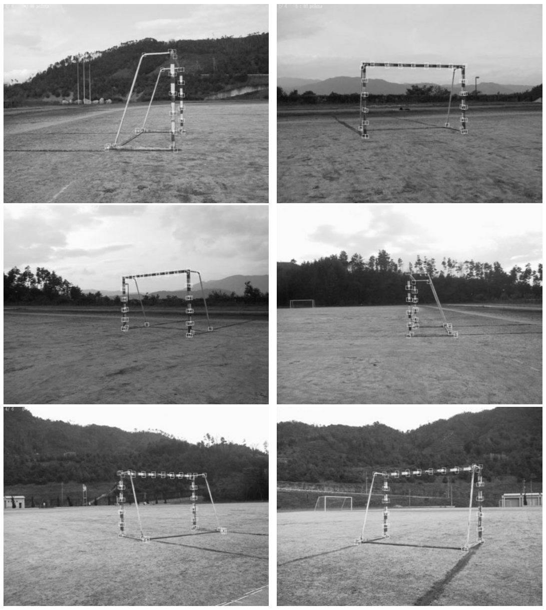

Figure 13: Images of the goal post. Angle information is obtained with the angle sensor simultaneously.

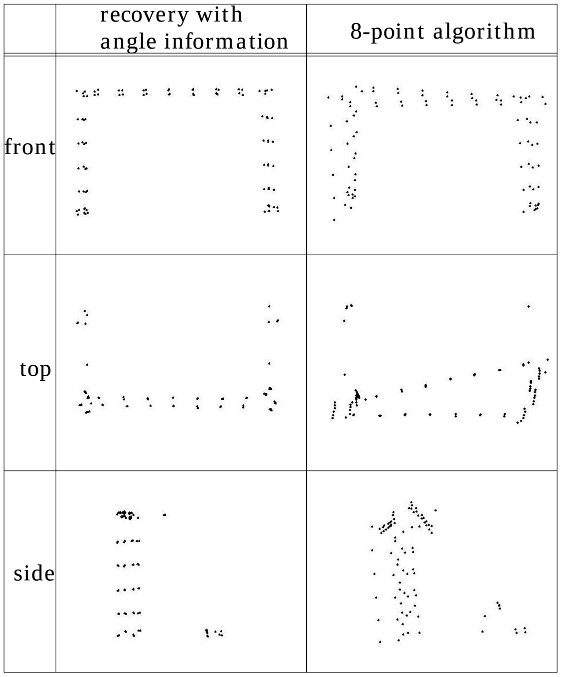

Figure 14: Resulting front, top, side views of the goal post with proposed method(left) and 8-point algorithm(right).

5.2 Experiment on Real Data

We next conducted the experiments on the real image and the angle data. To obtain angle information, we used 3 DOF angle sensor (MicroStrain Corp. 3DM–G) attached to the camera. Fig.13 shows all the images used for recovery. The size of the image and the focal length is same with the simulation experiments. The number of feature point is 88 each of which are tracked manually.

To compare with results without angle information, we applied 8 point algorithm[8] both with proposed method. Resulting front, top, side views are shown in Fig.14. As the feature points are selected along the straight lines, the recovered feature points should lie on the straight lines. In the top view and the side view of the recovered shape with proposed method, there exist position errors with the feature points, however, the error is very small that they can be improved with the bundle adjustment. On the other hand, resulting position of some feature points with the 8-point algorithm has large error which is difficult to improve with bundle adjustment. This shows the effectiveness of using angle information in the shape from motion problem.

6 Conclusions

The shape from motion problem is a difficult problem which cannot be solved in fast and stable algorithm generally. To cope with this problem, we proposed to incorporate the camera angle information. With this information, we can solve the problem only with linear algebra that leads to the fast and stable calculation. We also evaluated the effect of error of angle information where we could successfully obtain good results with error of 2 degrees.

References

[1] R. Szeliski and S. B. Kang: "Recovering 3D Shape and Motion from Image Streams using Non-Linear Least Squares," CVPR, 752–753, 1993.

[2] C. Tomasi and T. Kanade: "Shape and Motion from Image Streams under Orthography: a Factorization Method," IJCV 9(2), 137–154, 1992.

[3] J. Oliensis: "A Multi-Frame Structure-from-Motion Algorithm under Perspective Projection," IJCV 34(2/3), 163–192, 1999.

[4] A. Chiuso, R. Brockett and S. Soatto: "Optimal Structure from Motion: Local Ambiguities and Global Estimates," IJCV 39(3), 195–228, 2000.

[5] P. F. McLauchlan: "A Batch/Recursive Algorithm for 3D Scene Reconstruction," CVPR, 738–743, 2000.

[6] T. Jebara, A. Azarbayejani, and A. Pentland: "3D Structure from 2D Motion," IEEE Signal Processing Magazine, 66–84, 1999.

[7] D. J. Heeger, A. D. Jepson: "Subspace methods for recovering rigid motion I: Algorithm and Implementation," IJCV 7(2), 95–117, 1992.

[8] R. I. Hartley: "In Defence of the 8-point Algorithm," ICCV, 1064–1070, 1995.