8—8

Entire Building Shape Recovery from Near Distance Images

MIGITA Tsuyoshi*
Department of Intelligent Systems
Hiroshima City University

AMANO Akira†
Graduate School of Informatics
Kyoto University

ASADA Naoki‡
Department of Intelligent Systems
Hiroshima City University

Abstract

This paper describes a method to recover entire 3D shape of a building from an image sequence. This is a kind of "Shape and Motion recovery" problem, whereas conventional methods do not work well with images taken around a large object in near distance. Since such image sequence is a set of partial observations, 3D recovery becomes unstable. We first discuss the property of local minima of the nonlinear optimization function, and then describe a procedure to find the global minimum by avoiding pseudo solutions. Experiments using real images have shown that the proposed method successfully recovered 3D shapes from eleven sets of image sequence.

1 Introduction

This paper presents a method for recovering entire 3D building shape from an image sequence, based on the Shape and Motion recovery techniques. Images are taken at sufficient points to cover entire building, and are assumed to be taken at near distance from the building located in crowded urban area. In such a situation, each image contains partial observations of the building and the 3D shape recovery becomes difficult problem because there are many local minima in solving nonlinear equation.

Such difficulty is not considered and discussed fully in prior researches. Many conventional shape and motion recovery methods are proposed and they avoid this difficulty by assuming some restrictions to the images, implicitly or explicitly. In the Factorization method proposed by Tomasi and Kanade[1], images are assumed to be well approximated by linear projection, i.e. images are taken from far distance compared to the object size. Also in the building shape recovery method by Koch et.al.[2], camera trajectories are assumed to be relatively far from the objective buildings compared with its size, and hence each image contains almost all part of the building with relatively small perspective distortion.

*Address: 3-4-1 Ozuka-higashi, Asaminani-ku, Hiroshima, 731-3194 Japan. E-mail: migita@cv.its.hiroshima-cu.ac.jp

†Address: Yoshidahonmachi, Sakyoku, Kyoto, 606-8501 Japan. E-mail: amano@i.kyoto-u.ac.jp

‡E-mail: asada@its.hiroshima-cu.ac.jp

In the urban situation, however, buildings are so densely located that their images inevitably becomes close-ups, where each image contains limited part of the building. Thus the feature point correspondences become very sparse with the consequence that the objective equation has many local minima. To recover 3D shape and camera position from such image sequence, we need to deal with such local minima in the optimization process.

To recover entire shape of a building in realistic situation, we should deal with close-up images which have sparse feature point correspondences and have large perspective distortion. To cope with this problem, we propose incremental 3D shape recovery procedure. The key idea of our procedure is that we introduced trial-and-error search for our optimization process in order to find the optimal solution, instead of deterministic procedure. Note that, although our procedure automatically recovers 3D shapes with many image sets, some human controls are necessary to recover correct shape for difficult image sets.

In the following sections, we first formulate the problem, then propose procedure to avoid local minima. Finally, we present experimental results for several real image sets to show effectiveness of our method.

2 Entire Building Recovery

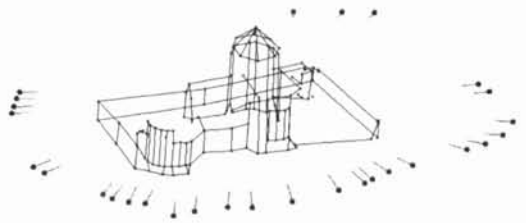
2.1 Formulation

Shape and motion recovery from an image sequence, is formulated as nonlinear least-squares problem[3] which minimizes the sum of squared reprojection errors. Specifically,

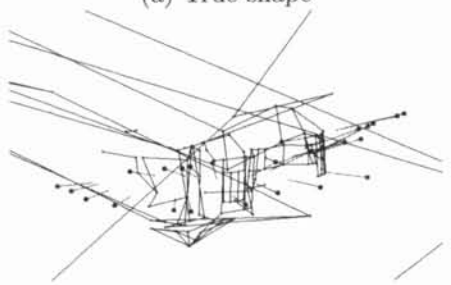
$$\arg \min_{\boldsymbol{s}_p \boldsymbol{t}_f R_f} \sum_{(f,p)\in S} |\tilde{\boldsymbol{u}}_{fp} - \mathcal{P}[R_f \boldsymbol{s}_p + \boldsymbol{t}_f]|^2 \qquad (1)$$

where $\boldsymbol{s}_p$ is the unknown 3D coordinates of p'th feature point, $R_f, \boldsymbol{t}_f$ are the unknown rotation and translation of f'th camera, $\tilde{\boldsymbol{u}}_{fp}$ is given 2D coordinates of p'th feature point in f'th image, S is the set of indices (f,p) over which the summation is calculated, and $\mathcal{P}$ denotes perspective projection.

Given initial value of the shape $\boldsymbol{s}_p$ and the camera position $R_f, \boldsymbol{t}_f$, optimal solution is calculated by nonlinear least-squares algorithm[3] such as Levenberg-Marquardt method or Preconditioned Conjugate Gradient method[4].



(a) True shape



(b) Recovered shape with naive optimization

Figure 1: Local minima of the problem.

2.2 Avoiding Local Minima

Since, there exist many local minima for this problem, occasionally, final estimation becomes one of them according to the initial value. Especially in the case of near-distance images, many local minima exists. For the feature points and camera positions which are illustrated in fig.1(a), recovered 3D shape and camera positions become as fig.1(b) when we use straight forward optimization to the whole dataset at once.

Problem difficulty can be characterized by amount of overlapping feature points among each image. This is evaluated by the ratio of size of S to $f \times p$. Hereafter, we refer this amount as *appearance ratio*. Experimentally, many local minima appears in optimization process if this ratio is low.

To avoid such local minima, we propose an incremental 3D shape recovery procedure. Denoting S_i as i'th subset of S, the procedure can be summarized as a searching process of S_i where optimal solution can be obtained for every S_i by using $R(S_{i-1})$ as initial value, where $R(S_{i-1})$ represents resulting solution of S_{i-1}. We start with some small set S_0 with which the associated shape and motion can be stably obtained without a priori initialization[3, 4]. Then gradually expand S_i until it gets to whole set S.

Occasionally, in the searching step, the situation that no suitable S_i can be found occurs. In such situation, we need to backtrack to prior step, and choose different S_{i-1}. This set expansion procedure is similar to Tomasi and Kanade's strategy[1], but the procedure of backtracking which is key part of local minima avoidance, is not considered in their method, and this is very important in nonlinear optimization process. Our goal is to find the path $\{S_0, S_1, ..., S_n\}$ such that the resulting optimal values $R(S_i)$ can be obtained by using $R(S_{i-1})$ as initial value. Hence obtained result $R(S_n)$ is global minimum of the underlying equation.

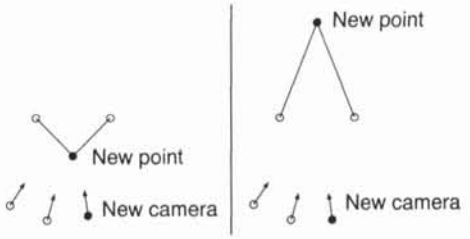


Figure 2: Point reversal.

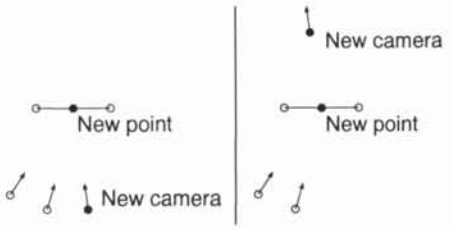


Figure 3: Camera reversal.

2.3 Analysis of Local Minima

In the optimization process, if resulting residue of equation(1) becomes relatively high, we can easily say that the considering set S_i is inappropriate and choose another set. However, in some situation, it is very difficult to choose another set with which the resulting shape becomes global optimal. In most case, they are classified into two following categories.

1. point reversal

   In fig. 2, the position of newly estimated point differs between left and right images. Feature point positions in the left are optimal, while positions in the right are local minimum. This is partial depth reversal problem. In prior Shape from Motion research, non-partial depth reversal problem is mentioned[3], where recovered depth of each point is coherently reversed. However, in our near-distance case, partial feature point reversal is observed.

2. camera reversal

   In fig. 3, the position of newly estimated camera differs. Camera positions in the left are optimal, while the right ones are local minimum. This is also connected to depth reversal problem. This happens due to low appearance ratio. If the appearance ratio is 100%, this situation will not occur.

In such case, reselection of S_i is necessary, while oridinary expansion is just used in usual case. In the next section, reselection procedures are described.

2.4 Optimization Procedure

Here, we describe the operations $S_{i-1} \stackrel{op}{\rightarrow} S_i$ which indicates the selection or reselection process of S_i from S_{i-1}. Although the selection of S_i is arbitrary, we used following selection scheme. S_i is

characterized by 3 variables, (s, e, L). The indices (f, p) in the set S_i satisfies following 2 conditions.

(a) The image number f should be in the range $s \leq f < e$.

(b) The point number p should appear at least L times in the set.

Mathematically, L in condition (b), must be larger than or equals to 2, because 3D position of a feature point observed in single image cannot be recovered by triangulation. Moreover, it is known that the 3D position calculated from only 2 images are unstable. To avoid these unstability, one should first recover with relatively large L which yields reliable estimation. Afterwards, decrease L gradually to 2 so that 3D position of every feature point is recovered. This control of L is also useful for local minima avoidance. If the estimation is considered to be a local minimum, increase of L might remove unstable feature points and/or camera.

Operations of $S_{i-1} \stackrel{op}{\rightarrow} S_i$ which we used is listed as follows.

1. Expand the set by increasing e.
2. Expand the set by decreasing s.
3. Expand the set by decreasing L.

On the other hand, to avoid local minima, operation for reselecting S_i is needed. Operations which we used is as follows.

1. Shrink the set by increasing L.

Our goal is to find the path S_i which yields true final estimation. For this purpose, a heuristic search method is employed which selects one operation at each step. Starting with S_{i-1}, first one of expansion operation is selected to make a candidate of S_i. If associated residue becomes larger than a threshold, another operation is selected to produce another candidate that gives better residue. If no improvement is made or no other operation is available, backtracking is performed to reselect S_{i-1}.

In the procedure, a local minima is detected by simple thresholding. However, this thresholding sometimes fails. In such case, manual control is necessary.

Also, selection of initial set S_0 is very important to achieve optimal solution. We used following criterions for selection of initial set.

- Appearance ratio is high.
- Relatively large L.
- If camera positions are roughly known, use images which have long baseline.
- Optimal solution can be calculated with the set.

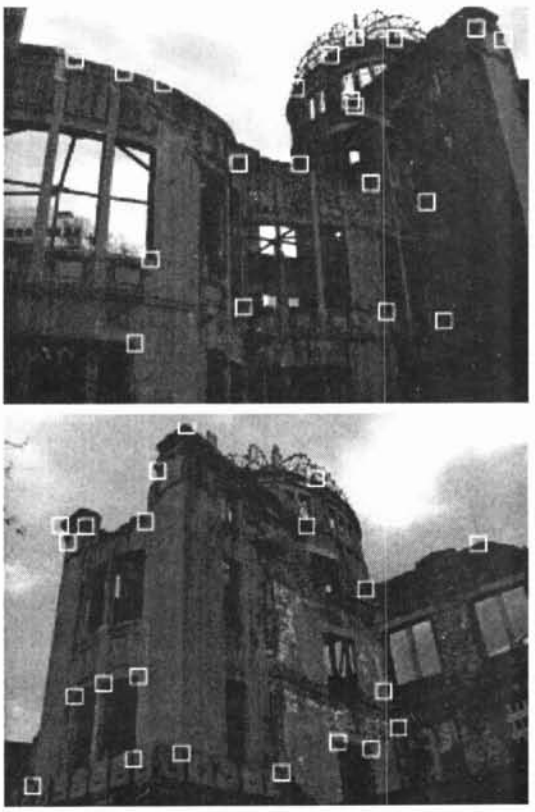


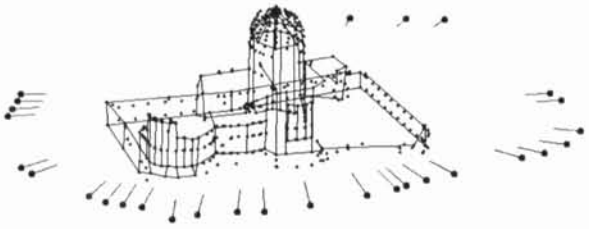
Figure 4: Two images of The Hiroshima Atomic Bomb Dome.

3 Experiments

We have applied our algorithm to 11 real image sets and successfully obtained 3D shapes and camera positions. Here, we show the results of two of these experiments: relatively easy one and relatively difficult one.

Fig. 4 shows 2 out of 29 images of the Hiroshima Atomic Bomb Dome, in which 2D coordinates of feature points are also shown. There are 469 feature points to represent the building shape, which are selected and matched manually. No image contains whole view of the building as shown in these pictures. Appearance ratio is 17%, which means it is relatively easy to recover 3D geometry. We start with 5 images and finally obtained 3D shape and camera positions shown in Fig.5. Four image sets including this set were able to recover their 3D shape and camera positions automatically.

Fig. 6 shows 4 out of 198 images of a gymnasium. There are 300 manually selected feature points to represent the building shape. Some images just contain limited part of the building as shown in figure. Appearance ratio is 5.1%. Thus, it is very difficult to recover 3D geometry. Left side of Fig. 7 shows initial estimation associated with the set S_0 which contains 20 images and 28 feature points. Right side of the Fig. 7 shows 3D shape and camera positions of intermediate step in search process. Many trial and backtracking were performed to produce final estimation shown in fig. 8,9. This is the entire building shape with positions and orientations of cameras displayed around it. From the recovered shape, we can say that rectangular shape and connecting an-



(a) Recovered 3D shape and camera positions.

(b) Side view of the dome with texture image.

Figure 5: Reconstructed Dome

gles of each building edges are well recovered. In this experiment, human decision of backtracking was occasionally needed as the problem is very difficult.

4 Conclusions

We proposed a method to recover entire 3D building shape from near-distance images which uses searching technique with backtrack to find optimal solution.

While this recovery problem is directly formulated as nonlinear optimization problem, we employed searching technique for finding optimal solution. The optimization process is formulated as path finding problem. Each node of path is the subset of indices over which the cost function is calculated. We employ heuristic search method to find path to avoid the local minima of the associated nonlinear cost function. When this heuristic search fails, one can manually control to search another path.

Experimental results on building image sets show the efficiency of our proposed method.

References

[1] C. Tomasi and T. Kanade: "Shape and Motion from Image Streams under Orthography: a Factorization Method," IJCV, Vol. 9, No. 2, pp. 137–154, 1992.

[2] R. Koch, M. Pollefeys and L. V. Gool: "Multi Viewpoint Stereo from Uncalibrated Video Sequences," ECCV '98, pp. I-55–71.

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

[4] A. Amano, T. Migita and N. Asada: "Stable Recovery of Shape and Motion from Partially Tracked Feature Points with Fast Nonlinear Optimization," Vision Interface 2002, pp. 244–251.

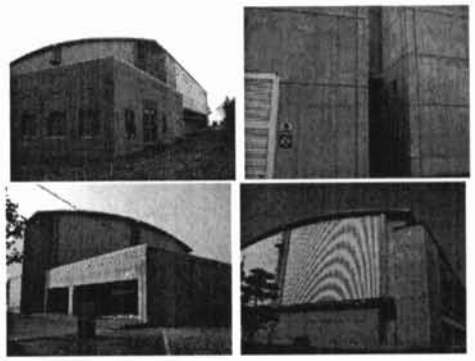


Figure 6: Images (4/198) of a gymnasium.

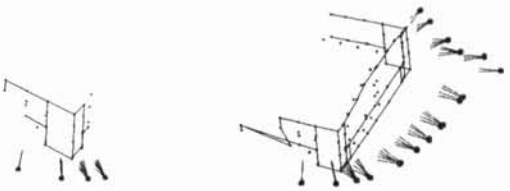


Figure 7: Initial(left) and intermediate(right) estimation.

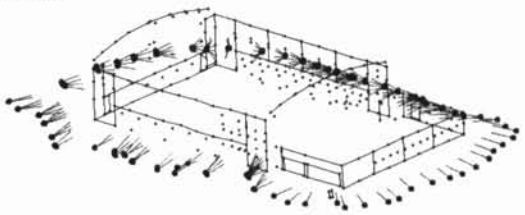


Figure 8: Recovered gymnasium and the cameras.

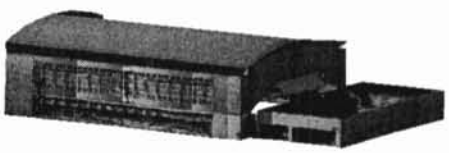


Figure 9: The recovered gymnasium with texture mapped.