INCREMENTAL PROCEDURE FOR RECOVERING ENTIRE BUILDING SHAPE FROM CLOSE RANGE IMAGES

MIGITA Tsuyoshi†, AMANO Akira‡, ASADA Naoki†
† Department of Intelligent Systems, Hiroshima City Univ.
Hiroshima, 731-3194, Japan. – {migita,asada}@cv.its.hiroshima-cu.ac.jp
‡Graduate School of Informatics, Kyoto Univ.
Kyoto, 606-8501, Japan. – amano@i.kyoto-u.ac.jp

KEY WORDS: Urban Building Reconstruction, Close Range Image, Computer Vision, Incremental Processing

ABSTRACT

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

1 INTRODUCTION

This paper presents a method for recovering entire 3D building shape from an close range image sequence, based on the Shape and Motion recovery techniques in the computer vision field. 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 problem becomes very difficult.

Such difficulty is not considered and discussed fully in prior computer vision researches. Many conventional shape and motion recovery methods are proposed, however, they exclude this difficulty from consideration by assuming some restrictions to the images, implicitly or explicitly.

In the Factorization method proposed by Tomasi and Kanade (Tomasi and Kanade, 1992), images are assumed to be well approximated by linear projection. To satisfy this approximation, all images should be taken from the far distance compared to the object size, and it is hard to realize this restriction in the entire building shape recovery at the urban situation. Also in the building shape recovery method by Koch et.al. (R. Koch and Gool, 1998), the 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.

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 which naturally becomes nonlinear, has many local minima. To recover the 3D shape and the camera positions 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 SHAPE RECOVERY

2.1 Formulation

Shape and motion recovery from an image sequence, is formulated as nonlinear least-squares problem (Szeliski and Kang, 1993) which minimizes the sum of squared reprojection errors. Specifically,

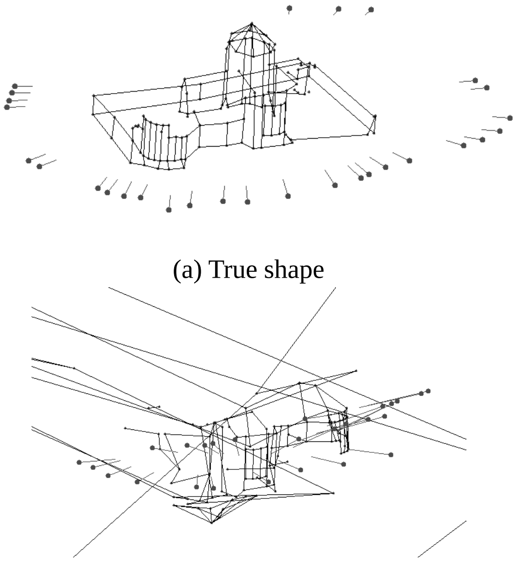
$$\arg\min_{\boldsymbol{s}_p \boldsymbol{t}_f R_f} \sum_{(f,p)\in S} |\tilde{\boldsymbol{u}}_{fp} - \mathcal{P}[R_f \boldsymbol{s}_p + \boldsymbol{t}_f]|^2 \qquad (1)$$

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(Szeliski and Kang, 1993) such as Levenberg-Marquardt method or Preconditioned Conjugate Gradient method(Amano et al., 2002).

2.2 Avoiding Local Minima

Although many researchers have considered this Structure and Motion recovery problem, they usually deal with



(a) True shape

(b) Recovered shape with naive optimization

Figure 1: Local minima of the problem.

somewhat restricted dataset to reduce difficulty involving local minima, many of which, however, instrinsically arise in the problem with general configuration. Qualitatively, number of local minima decreases when the distance from the camera to the object compared to the object size becomes large, and the variation of the distance becomes small, and the number of the feature points in each image becomes large which has major effect compared to the others. In the previous research, two or threee of these restrictions are assumed implicitly or explicitly.

Considering entire building shape recovery in the urban situation, it is hard to satisfy these restrictions. Especially in the case of close range images, many local minima exists. With the data set of the feature points and the camera positions shown in fig.1(a), recovered 3D shape and camera positions become as fig.1(b) when we use a straight forward optimization to the whole dataset at once.

The problem's difficulty can be characterized by the 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 to this amount as *appearance ratio*. Experimentally, many local minima appears in the optimization process if this ratio is low.

Here, we consider the most general situation; the distance between the building and the camera is small, the variation of the distance is large, and the appearance ratio is low. To avoid local minima described above, 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 the resulting solution of the S_{i-1}. We start with some small set S_0 with which the associate shape and motion can be stably obtained without a priori initialization(Szeliski and Kang, 1993, Amano et al., 2002). Then gradually expand the S_i until it gets to whole set S.

In the searching step, we occasionally meet the situation where no suitable S_i can be found. In such situation, we need to backtrack to the prior step, and choose different S_{i-1}. This set expansion procedure is similar to the Tomasi and Kanade's strategy(Tomasi and Kanade, 1992), but the procedure of backtracking which is the key part of the local minima avoidance, is not considered in their method, and this is very important in the 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 the initial values. Hence obtained result $R(S_n)$ becomes the global minimum of the underlying equation.

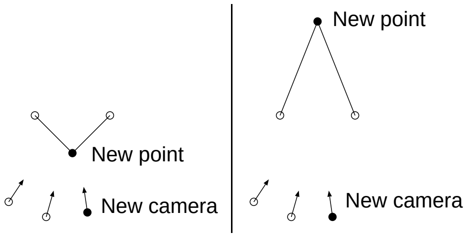


Figure 2: Point reversal.

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. Two typical situations of such case can be described as follows.

1. point reversal

   In Figure 2, the position of newly estimated point (black dot) differs between left and right case. The feature point positions in the left are optimal, while that in the right are local minimum. This is partial depth reversal problem. In the prior Shape from Motion researches, the complete depth reversal problem is mentioned(Szeliski and Kang, 1993), where recovered depth of each point is coherently reversed. However, in the close range image case, partial feature point reversal is observed.

2. camera reversal

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

When these case occurs, reselection of S_i is necessary, while the ordinary expansion is just used in the usual case. In the next section, reselection procedures are described.

2.4 Path Finding Procedure

Here, we describe the operations $S_{i-1} \overset{op}{\rightarrow} S_i$ which indicates the selection or reselection process of S_i from S_{i-1}.

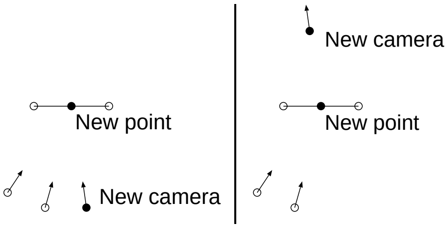


Figure 3: Camera reversal.

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 instability, 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 incorrectly estimated feature points and/or camera.

Operations of $S_{i-1} \overset{op}{\rightarrow} S_i$ towards the entire set are listed as follows.

E1 Expand the set by increasing e.

E2 Expand the set by decreasing s.

E3 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.

S1 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}.

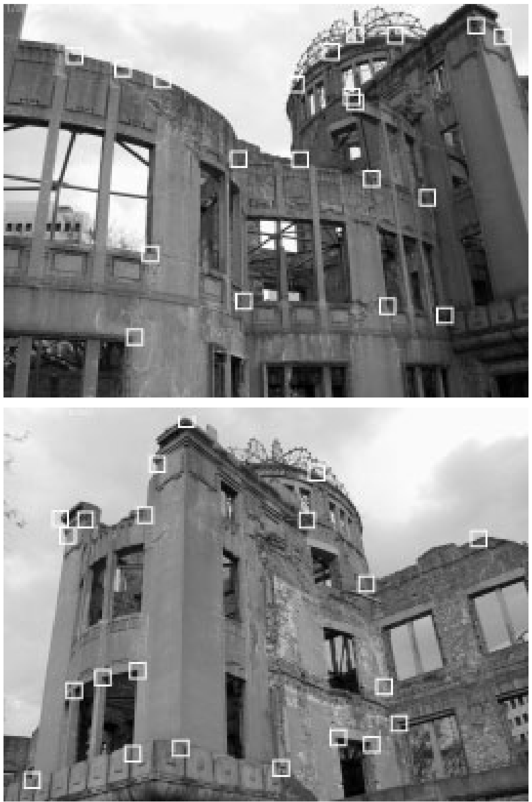


Figure 4: Two images of The Hiroshima Atomic Bomb Dome.

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

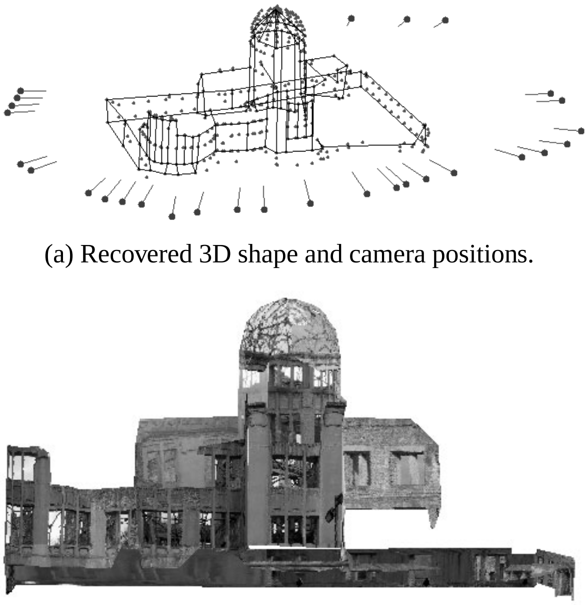
Also, selection of initial set S_0 is very important to achieve optimal solution. We used following criteria for selection of initial set.

- 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.

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



(a) Recovered 3D shape and camera positions.

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

Figure 5: Reconstructed Dome

and camera positions shown in Fig.5. Four image sets including this one 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 Figure 8 where the entire building shape is displayed with positions and orientations of cameras around it, and in Figure 9, recovered shape is shown with texture mapped. From the recovered shape, we can say that rectangular shape and connecting angles 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 close range images. The recovery problem is directly formulated as nonlinear optimization problem, and we proposed the incremental procedure for finding optimal solution. The optimization process is formulated as path finding problem. Each node of a path is the subset of indices over which the cost function is calculated. We employ heuristic search method to find a path to avoid the local minima of the associated nonlinear cost function. When this heuristic search fails, one can manually control the system to search another path.

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

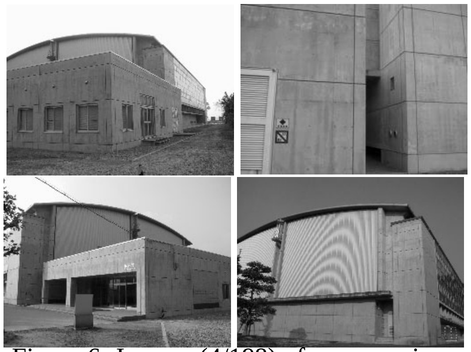


Figure 6: Images (4/198) of a gymnasium.

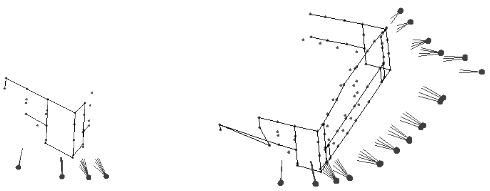


Figure 7: Initial(left) and intermediate(right) estimation.

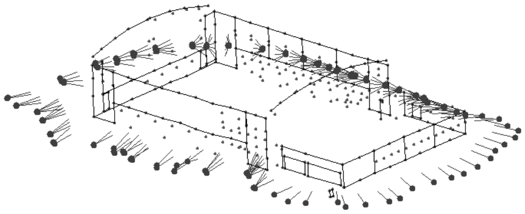


Figure 8: Recovered gymnasium and the cameras.

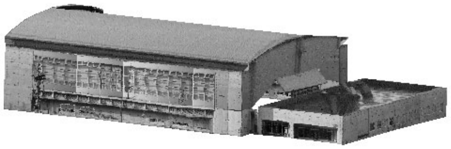


Figure 9: The recovered gymnasium with texture mapped.

When applying our method to photogrammetry, evaluating and guaranteeing the precision of recovered shape is important. They are left as future works.

REFERENCES

Amano, A., Migita, T. and Asada, N., 2002. Stable recovery of shape and motion from partially tracked feature points with fast nonlinear optimization. Vision Interface 2002 pp. 244–251.

Koch, R., Pollefeys, M. and Gool, L. V., 1998. Multi viewpoint stereo from uncalibrated video sequences. ECCV ’98 pp. I–55–71.

Szeliski, R. and Kang, S. B., 1993. Recovering 3d shape and motion from image streams using non-linear least squares. CVPR pp. 752–753.

Tomasi, C. and Kanade, T., 1992. Shape and motion from image streams under orthography: a factorization method. IJCV 9, pp. 137–154.