Stable Recovery of Shape and Motion from Partially Tracked Feature Points with Fast Nonlinear Optimization

Akira Amano, Tsuyoshi Migita, Naoki Asada
Faculty of Information Sciences, Hiroshima City University
{a-amano,migita,asada}@cv.its.hiroshima-cu.ac.jp

Abstract

The linearized approach to the shape from motion problem, e.g. the factorization method, is robust to find a unique solution with fast computation, but the occlusion and perspective distortion are out of scope in the linear formulation. In contrast, the nonlinear approach is free from such limitations, yet it involves two problems; one is to find the globally optimal solution and the other to reduce the computation time.

In this paper, we present an effective nonlinear optimization method to recover 3D shape and motion. To overcome the shortcomings of the nonlinear scheme, we propose "double search (DS) procedure" and "preconditioned conjugate gradient (PCG) algorithm". The DS procedure enables us to find two major solutions that correspond to the true and false shapes and then we select the globally optimal solution by evaluating the error of them. The PCG algorithm is an improved CG one whose computational performance is several times faster than the conventional Levenberg-Marquardt (LM) algorithm.

We carried out experiments with simulation and real data, and the results have demonstrated that the proposed method allows us to obtain the correct shape and motion with 3-9 times faster computation than the LM algorithm. Finally we have shown the applicability of the method to a large building reconstruction from a set of partially tracked feature points.

1 Introduction

The common interest in computer vision is to obtain 3D information from 2D images. Various approaches have been studied so far, and one practical method among them is "Shape from Motion"; that is, 3D shape of an object and relative motion between the object and camera are recovered from an image sequence.

3D reconstruction is the inverse problem of imaging process by perspective projection, and thus the relationship between 3D coordinates and the corresponding 2D ones is represented by nonlinear expressions in terms of the depth parameter. There are two different mathematical schemes to solve the nonlinear problem; one is linear approximation and the other nonlinear optimization.

The linearized approach to the shape from motion, e.g. the factorization method, is indeed robust to find a unique solution with fast computation, but some problems such as occlusion and perspective distortion are out of scope due to the linear formulation. In contrast, the nonlinear approach is free from such limitations, yet it involves two problems; one is to find the globally optimal solution and the other to reduce the computation time of the iterative search process.

In this paper, we propose a nonlinear optimization method to recover 3D shape and motion with fast and stable computation. To overcome the shortcomings of the nonlinear scheme described above, we have developed "double search (DS) procedure" and "preconditioned conjugate gradient (PCG) algorithm". The DS procedure enables us to find two major solutions that correspond to the true and false shapes and then we can select the globally optimal solution by evaluating the error of them. The PCG algorithm is an improved CG one whose computation performance is several times faster than the conventional Levenberg-Marquardt algorithm.

In what follows, we first give a brief survey on the shape from motion problem, then formalize it in the scheme of nonlinear optimization. Next, the DS procedure and PCG method are presented with technical details. The performance of the proposed method is evaluated quantitatively by using simulation and real data. Finally, we will show some experimental results of reconstructing buildings from a set of partially observed image data.

2 Shape from Motion

In shape from motion studies, "eight point algorithm[1]" established the theoretical framework of 3D information recovery from two images. The method however hardly works in actual situations because it takes no account of the observation error. Therefore the method is not robust to feature

point noises, and the resultant shapes are often seriously distorted.

A straightforward way to enhance the stability of the shape from motion method is the use of a large number of images as redundant information with least squares criterion. This scheme is effective in many cases, but it increases the computational cost exponentially, especially in the large scale nonlinear problem. This is why many researchers believe that the linear approximation is the feasible approach to the shape from motion problem.

The factorization method proposed by Tomasi and Kanade[4] is the most successful method that realizes the fast and stable recovery of 3D shape and motion by using linear algebra with singular value decomposition (SVD). However, such a linearized formulation has intrinsic limitations to express the occluded feature points and the perspective distortion. A lot of efforts have been made on the original factorization to deal with partially tracked feature points[4] and perspective projection[5]. Note that such extensions not only increase the computational cost but also lose the merit of stable computation.

Another approach to the shape from motion is based on the nonlinear computation scheme that involves the following advantages. First, the perspective projection is represented properly in the formulation without any approximation. This means that 3D shape is recovered correctly from severely distorted images by perspective projection. Second, 3D reconstruction is performed from even partially tracked feature points in the image sequence. This means that the optimization algorithm accepts any set of feature points whether occlusion occurs or not during the observation.

Although the nonlinear approach has the great advantages described above, we have to consider the following issues when using it.

- How to find the globally optimal solution.

  In nonlinear optimization, the most important argument is to determine an appropriate initial value based on the optimization algorithm. If we begin the computation with a good starting value, we can reach the globally minimal solution in the energy space of the evaluation function; otherwise we fall into a local minimum that is a false solution. Therefore the initial value selection according to the solution search algorithm is the most significant issue. Surprisingly, there are a few discussion on this matter. Weng et.al.[8] claim the use of the solution by linear algorithm as a initial value of the nonlinear optimization. However, they employed the epipolar based linear algorithm using only two images, so that the initial value is not reliable and of course the occlusion is not considered.

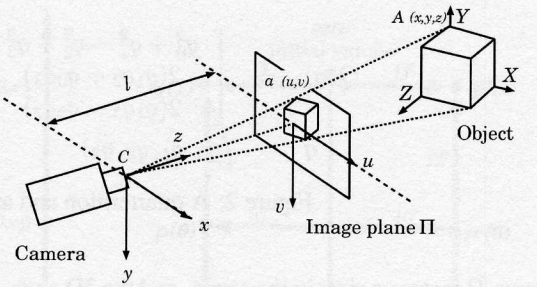


Figure 1: Projection model.

- How to reduce the computation time.

  Nonlinear optimization usually takes large computation time in the iterative search process. Therefore, we need a fast algorithm applicable to the large scale computation like shape from motion problem. The Levenberg-Marquardt (LM in short) algorithm is widely used by many researches[7][8], and Szeliski and Kang[6] employed it to the problem size of 100 frames. Recently McLauchlan[9] has proved that nonlinear optimization is a practical method to solve the shape from motion problem by using a faster LM algorithm which takes 100 to 200 times long computation compared with linear method.

3 Problem Definition and Formulation

Shape and motion parameters are represented nonlinearly as the trajectories of the projected feature points by perspective projection formula. Therefore we can obtain shape and motion by calculating this inverse problem with nonlinear optimization.

While the projection formula are common, there are various formulation in terms of the variable set selection. Szeliski and Kang[6] have shown a simple formulation, yet it needs some modification to fit each problem settings and the formulation is not discussed in detail.

In this section, we present the definition of the problem and its formulation.

3.1 Shape and motion in perspective projection

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

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

$$Q(\boldsymbol{q}) = \begin{bmatrix} q_0^2+q_1^2-q_2^2-q_3^2 & 2(q_1q_2-q_0q_3) & 2(q_1q_3+q_0q_2) \\ 2(q_1q_2+q_0q_3) & q_0^2-q_1^2+q_2^2-q_3^2 & 2(q_2q_3-q_0q_1) \\ 2(q_1q_3-q_0q_2) & 2(q_2q_3+q_0q_1) & q_0^2-q_1^2-q_2^2+q_3^2 \end{bmatrix}$$
$$\boldsymbol{q} = (q_0, q_1, q_2, q_3)$$

Figure 2: A quarternion and associated scaled-rotation matrix.

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 fth frame rotation of object coordinate system viewed from 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 fth frame translation of object coordinate system viewed from camera coordinate system.

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

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

When pth feature point at fth frame $\boldsymbol{s}_{fp}$ is projected to the image(Fig.1), 2D coordinate $\boldsymbol{u}_{fp} = (u_{fp}, v_{fp})^T$ becomes as follows,

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

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

$$\mathcal{P}\begin{bmatrix} x \\ y \\ z \end{bmatrix} := \frac{l}{z}\begin{bmatrix} x \\ y \end{bmatrix}$$

Among several representation of rotation matrix, the quarternion is used in shape from motion problem because the differential calculation is fast. Note that we need not use unit quarternion because $\mathcal{P}(\boldsymbol{s}) = \mathcal{P}(k\boldsymbol{s})$ stands for arbitrary scale factor $k \neq 0$.

When we define scale factor $k_f = l/t_z$, Eq.(1) becomes as follows.

$$\boldsymbol{u}_{fp} = \mathcal{P}\left[(k_f R_f)\boldsymbol{s}_p + (k_f \boldsymbol{t}_f)\right] \quad (2)$$
$$= \mathcal{P}\left(Q(\boldsymbol{q}_f)\boldsymbol{s}_p + \begin{bmatrix} X_f \\ Y_f \\ l \end{bmatrix}\right) \quad (3)$$
$$= \frac{l}{Q_{f3}\boldsymbol{s}_p + l}\begin{bmatrix} Q_{f1}\boldsymbol{s}_p + X_f \\ Q_{f2}\boldsymbol{s}_p + Y_f \end{bmatrix} \quad (4)$$
$$= \frac{1}{l^{-1}Q_{f3}\boldsymbol{s}_p + 1}\begin{bmatrix} Q_{f1}\boldsymbol{s}_p + X_f \\ Q_{f2}\boldsymbol{s}_p + Y_f \end{bmatrix} \quad (5)$$

Here, rotation matrix with scale factor $Q(\boldsymbol{q}_f)$ is defined as Fig.2 using quarternion $\boldsymbol{q}$. Note that ith row vector of $Q(q_f)$ is denoted as Q_{fi}.

We introduce projective distortion factor λ to l^{-1}.

$$\boldsymbol{u}_{fp} = \frac{1}{1 + (\lambda l^{-1})Q_{f3}\boldsymbol{s}_p}\begin{bmatrix} Q_{f1}\boldsymbol{s}_p + X_f \\ Q_{f2}\boldsymbol{s}_p + Y_f \end{bmatrix}. \quad (6)$$

For the perspective projection, the equation of projection is denoted with $\lambda = 1$, and in scaled-orthographic projection, equation is denoted with $\lambda = 0$.

3.2 Nonlinear optimization

When we represent the observed feature point coordinates by $\tilde{\boldsymbol{u}}_{fp} = (\tilde{u}_{fp}, \tilde{v}_{fp})$, shape from motion problem is formulated as optimization of following error function.

$$\min_{\boldsymbol{s}_p, R_f, \boldsymbol{t}_f} \frac{1}{FP}\sum_{(f,p)} ||\tilde{\boldsymbol{u}}_{fp} - \boldsymbol{u}_{fp}||^2 \quad (7)$$

This includes $N = 3P + 6F$ unknown variables which are denoted by a vector representation $\boldsymbol{x}$ as follows.

$$\boldsymbol{x} = (x_1, y_1, z_1, \cdots, x_P, y_P, z_P, X_1, q_{10}, q_{11}, q_{12}, q_{13}, Y_1, \cdots, X_F, q_{F0}, q_{F1}, q_{F2}, q_{F3}, Y_F)^T \quad (8)$$

Also we show the difference between every $\boldsymbol{u}_{fp}$ and $\tilde{\boldsymbol{u}}_{fp}$ by a vector $\boldsymbol{r}$ as follows.

$$\boldsymbol{r} = (g_{111}, g_{112}, \cdots, g_{FP2})^T \quad (9)$$
$$\begin{bmatrix} g_{fp1} \\ g_{fp2} \end{bmatrix} = \begin{bmatrix} u_{fp} \\ v_{fp} \end{bmatrix} - \begin{bmatrix} \tilde{u}_{fp} \\ \tilde{v}_{fp} \end{bmatrix} \quad (10)$$

With these variables, we have the cost function of Eq.(7) as follows.

$$E(\boldsymbol{x}) = \frac{1}{FP}\boldsymbol{r}^T\boldsymbol{r}, \quad (11)$$

The optimal solution is shown by $\boldsymbol{p}(\lambda)$ which is as a function of λ.

It is obvious that the globally optimal solution of this system is not unique because the cost value is invariant on translation of world origin and arbitrary rotation and scaling.

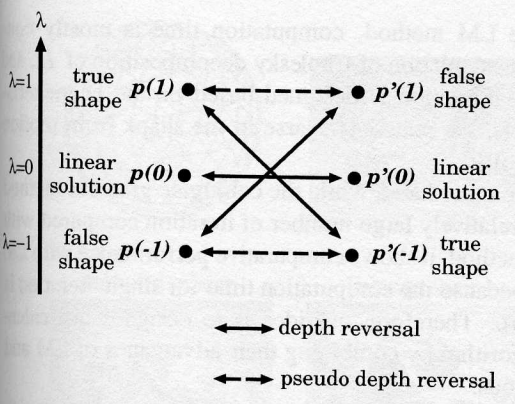


Figure 3: Two major local minima.

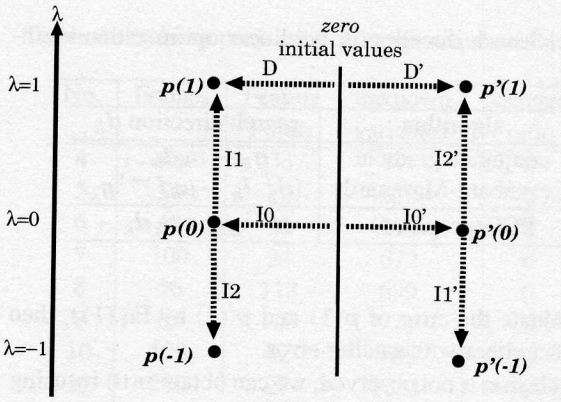


Figure 4: Double search procedure.

4 Double Search Procedure

In shape from motion problem, there exist two major solutions; that is, there are dual interpretations of the reconstructed shape with respect to the depth under the condition of linear projection. This kind of ambiguity exists even under the perspective projection[2][3]. One is globally optimal solution which corresponds to the correct shape, and the other is semi-optimal solution which is known as depth reversal shape or Necker's reversal.

4.1 Avoidance of local minima

The issue on the local minima avoidance is seldom discussed so far. From our experiments, the more occluded feature points are included the more local minima appear. However, there exist only two major optimal solutions if a small number of feature points are occluded. One is globally optimal solution which corresponds to the true shape, and the other is semi-optimal solution which is known as a depth reversal shape or Necker's reversal. In this paper, we discuss the latter situation where the small occlusion is observed.

In order to avoid the false solution, the following method is widely used[6]; that is, the shape of the first stage computation is reversed in terms of depth and it is used as the initial value for the second stage computation. However, we can not always obtain the correct shape with this method because the depth reversed shape becomes seriously distorted when the images are strongly distorted by perspective projection.

4.2 Properties of major solutions

When we recover the shape under the scaled-orthographic projection, we obtain two major minima which are pure depth reversal to each other. Under perspective projection, we obtain two corresponding local minima which are pseudo depth reversal to each other. One of the shape corresponds to global minimum which we call true shape, and the other is just a local minimum which we call false shape.

By using λ in Eq.(6), we can express these shapes as $\boldsymbol{p}(\lambda)$. Under scaled-orthographic projection, λ becomes 0, thus two shapes can be denoted as $\boldsymbol{p}(0)$ and $\boldsymbol{p}'(0)$ which are depth reversal to each other(Fig.3).

When we set λ to 1, true shape becomes $\boldsymbol{p}(1)$. From Eq.(6), 2D projected coordinates of a feature point on the screen becomes equal to the point whose z_p value is depth reversed to $-z_p$ and λ is set to -1. As we represent the depth reversed shape of $\boldsymbol{p}(\lambda)$ as $\boldsymbol{p}'(-\lambda)$, we can say that $\boldsymbol{p}(1)$ and $\boldsymbol{p}'(-1)$ are equal shape. However, $\boldsymbol{p}(1)$ and $\boldsymbol{p}'(1)$ are not depth reversed shape. We call these relation "pseudo depth reversal".

4.3 Double search procedure

As the resultant shape of $\boldsymbol{p}'(1)$ becomes strongly distorted when the perspective distortion is severe, we use the resultant shape $\boldsymbol{p}(0)$ obtained under linear projection model as initial value of nonlinear optimization. As we easily obtain $\boldsymbol{p}'(0)$ by reversing depth information of $\boldsymbol{p}(0)$, starting from these two shapes, we obtain $\boldsymbol{p}(1)$ and $\boldsymbol{p}(-1)$ by gradually varying λ from 0 to ± 1. As $\boldsymbol{p}'(1)$ is equivalent to $\boldsymbol{p}(-1)$ except for depth direction, we obtain $\boldsymbol{p}'(1)$ by reversing depth information of $\boldsymbol{p}(-1)$. We call this "**Double Search**" procedure (Fig.4).

The procedure is summarized as follows:

1. Obtain a linear solution $\boldsymbol{p}(0)$ (I0, I0' in Fig.4).
2. Change $\lambda = 0 \rightarrow 1$ in $\boldsymbol{p}(\lambda)$ continuously and have $\boldsymbol{p}(1)$. As the same manner, change $\lambda = 0 \rightarrow -1$ in $\boldsymbol{p}(\lambda)$ continuously and have $\boldsymbol{p}(-1)$ (I1, I2 or I1', I2' in Fig.4).
3. Reverse the depth of $\boldsymbol{p}(-1)$ and obtain $\boldsymbol{p}'(1)$.

Table 1: Search direction in nonlinear optimization methods.

algorithm	search direction $\boldsymbol{d}_k$
conjugate gradient	$\boldsymbol{g}_k - \beta_k \boldsymbol{d}_{k-1}$
Levenberg-Marquardt	$(J_k^T J_k + \mu_k I)^{-1} \boldsymbol{g}_k$
PCG(proposed)	$\hat{H}^{-1} \boldsymbol{g}_k - \beta_k \boldsymbol{d}_{k-1}$

4. Evaluate the error of $\boldsymbol{p}(1)$ and $\boldsymbol{p}'(1)$ by Eq.(11), then select either with smaller error.

If occlusion is not observed, we can obtain $\boldsymbol{p}(0)$ by using the factorization method.

5 Fast Optimization Algorithm: PCG

To obtain the optimal solution for Eq.(11), gradient algorithm such as Levenberg-Marquardt method[10, 11] is widely used[6]. In the gradient algorithm, minimization of $E(\boldsymbol{x})$ is performed by iteration of the following formula,

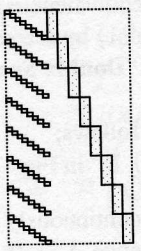
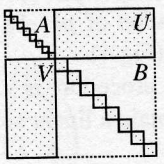
$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \alpha \boldsymbol{d}_k, \quad \alpha = \arg\min_{\alpha} E(\boldsymbol{x}_k - \alpha \boldsymbol{d}_k)$$

where, search direction $\boldsymbol{d}_k$ is calculated using gradient vector $\boldsymbol{g}_k := \nabla E(\boldsymbol{x}_k)$ when $\boldsymbol{x}$ is estimated as $\boldsymbol{x}_k$. Major algorithm for nonlinear optimization is shown in Table 1 where $\hat{H}_k$ denotes approximated Hessian matrix, J_k denotes Jacobian matrix and β_k, μ_k are parameters.

In the LM method, Hessian matrix is approximated by $(J_k^T J_k + \mu_k I)$ where Jacobian J is defined as follows.

$$J = \begin{bmatrix} \frac{\partial g_{111}}{\partial x_1} & \frac{\partial g_{111}}{\partial x_2} & \cdots & \frac{\partial g_{111}}{\partial x_N} \\ \frac{\partial g_{112}}{\partial x_1} & \frac{\partial g_{112}}{\partial x_2} & \cdots & \frac{\partial g_{112}}{\partial x_N} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial g_{FP2}}{\partial x_1} & \frac{\partial g_{FP2}}{\partial x_2} & \cdots & \frac{\partial g_{FP2}}{\partial x_N} \end{bmatrix},$$

In the shape from motion problem, the form of J and approximated Hessian $\hat{H}$ generally becomes as follows,

$J =$  , $\hat{H} =$ 

where white space represents 0 values. In $\hat{H}$, blocks A and B are block diagonal and the size of each block is 3 in A, 6 in B.

In the LM method, computation time is mostly consumed on calculation of Cholesky decomposition of $\hat{H}$, and the faster algorithm is designed based on the characteristics; that is, the matrix is sparse in the shape from motion problem[6][9].

On the other hand, while the conjugate gradient method requires relatively large number of iteration compared with the LM method, it shows comparative performance with LM method because the computation time for single iteration is very short. Therefore, our idea is to design a fast calculation algorithm by combining their advantages of LM and CG methods.

Here, we describe the "Preconditioned Conjugate Gradient(PCG) method" applied for nonlinear optimization. Note that PCG method is used in linear optimization problems in the past. We use $\hat{H}$ with U and V part filled with 0 as preconditioning matrix C. As C becomes block diagonal matrix of width 9, Cholesky decomposition can be performed with $O(N)$ computation time. Note that this approximation does not affect the accuracy of the final result of optimization. The algorithm is denoted as follows.

Suppose a current estimation of $\boldsymbol{x}$ as $\boldsymbol{x}_k$, first search direction is calculated as $\boldsymbol{d}_1 = C^{-1}\boldsymbol{g}_1$, afterwards, estimation of $\boldsymbol{x}$ is calculated as follows.

$$\begin{cases} \boldsymbol{d}_k &= C^{-1}\boldsymbol{g}_k + \beta \boldsymbol{d}_{k-1} \\ \boldsymbol{x}_{k+1} &= \boldsymbol{x}_k - \alpha \boldsymbol{d}_k \end{cases}$$

where,

$$\begin{cases} \alpha &= \arg\min_{\alpha} E(\boldsymbol{x}_k - \alpha \boldsymbol{d}_k) \\ \beta &= (C^{-1}\boldsymbol{g}_k)^T(\boldsymbol{g}_k - \boldsymbol{g}_{k-1}) / \boldsymbol{g}_{k-1}^T C^{-1} \boldsymbol{g}_{k-1} \end{cases}$$

Note that C should be recalculated after several iteration. In our experiment, C is recalculated after every 16 iterations.

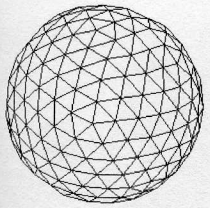
Experimentally, this PCG method is 3 to 9 times faster than the LM method, so that it is efficient enough to use in real applications.

6 Experiments and Evaluation

Experimental results on shape and motion recovery with nonlinear optimization method are described here to show that our method is fast and robust.

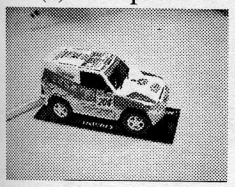
First, simulation data of hemisphere whose diameter is 200 pixel is used to recover the 3D shape (Fig.5(a)). 90 frames are generated by rotating it in 90 degrees, and it is viewed from 250 pixel far. The images are strongly distorted by perspective projection in this situation, and that 20% of the feature points were occluded.

We applied the PCG method to obtain $\boldsymbol{p}(0)$ with $\lambda = 0$ which means that the linear projection problem is solved by nonlinear optimization, then used the double search procedure to obtain $\boldsymbol{p}(1)$. In this case, we successfully obtain the



(a)hemisphere

(b)sequence 1



(b)sequence 2

(c)sequence 3

Figure 5: First frames of each image sequence.

Table 4: Problem properties of image sequences 4 – 14.

No.	Frames	Feature Points	unknown variables	occlusion ratio
4	63	33	477	0
5	50	70	510	0
6	65	92	666	0
7	100	24	672	0
8	56	118	690	0
9	67	123	771	0
10	100	62	786	0
11	240	27	1521	0
12	24	125	529	44%
13	65	125	765	5%
14	65	130	780	9%

correct shape. Note that the naive linear method cannot be applied to this sequence because occluded feature points are included. The resultant 3D accuracy was 0.0 in pixel unit for perspective projection result $\boldsymbol{p}(1)$, and 17.9 for linear projection result $\boldsymbol{p}(0)$. This demonstrates the accuracy of recovered shape using nonlinear method which represents the perspective projection without any approximation.

A car model ($65 \times 70 \times 155$ mm) is used to take three different image sequences whose problem properties are shown in Table 2: number of frames F, number of feature points P, number of variables, distance between object and camera, and the object extent in degree. First image of each sequence is shown in Fig.5(b)–(d).

The double search procedure is applied to those image sequences, and we successfully obtained the correct shape for all sequences. Note that the conventional search method to calculate $\boldsymbol{p}(1)$ immediately does not work in these cases. Table 3 shows the resultant residues, computation times and number of iterations. We used a PC with PentiumIII 850 MHz for the performance test. From the results of computation times, we find that the PCG method is 3 to 9 times faster than LM method to obtain 3D shape and motion with the same accuracy. Since the residues of Eq.(11) show the re-projection error of recovered shape that means difference between 2D coordinates of given feature points and projected 2D coordinates of recovered shape, it indicates the accuracy of recovered shape. From Table 3, it is shown that both residues of nonlinear and linear(factorization) methods are small in the image sequence 1 which includes less perspective distortion. However, the residues of the linear method becomes larger when the perspective distortion is not ignored in the image sequence 3. These results demonstrates the effectiveness of the nonlinear method which deals with the perspective projection properly.

We prepared the other real image sequences from 4 to 12 of a rotating cube($80 \times 80 \times 80$ mm). Note that the sequences from 4 to 11 are free from occlusion, but the sequence 12 has occlusion in its feature points. The sequence 13 and 14 are the scene of a car model with a little occlusion. Table 4 shows these properties of image sequences. The initial value $\boldsymbol{p}(0)$ is obtained by the PCG method, and we successfully obtained the correct shape and motion by the double search method for both LM and PCG methods. The comparison of the computation time between LM and PCG methods is shown in Fig.6. This shows that the performance of PCG method is 3 to 9 times faster than that of LM method.

Finally, we have carried out an experiment to recover some large objects from partially tracked feature points: the Hiroshima Atomic Bomb Dome and a stadium. The first and 15th image from 29 image with feature points are shown in Fig.7(a),(b). We have 469 feature points detected from the image sequence, and the reconstructed 3D points and camera positions are successfully recovered as shown in Fig.7(c). Fig.7(d) is a side view of the dome with texture image. Another experiment on building recovery is shown in Fig.8. (a) shows the first image of stadium out of 144 images, and 239 feature points for the sequence are used to recover it. (b) shows the top view of recovered stadium with recovered camera positions. From these results, we have confirmed that the proposed method works successfully for such a large buildings from even partially tracked feature points.

Table 2: Problem properties of image sequences 1 – 3.

	sequence 1	sequence 2	sequence 3
P(feature point) × F(frame)	90×81	187×19	269×16
variables	756	675	903
object distance (mm)	1100	300	150
object extent (deg)	8	25	40

Table 3: Comparison of computation time.

	sequence 1			sequence 2			sequence 3		
	time(sec)	iter.	residue	time(sec)	iter.	residue	time(sec)	iter.	residue
PCG	0.38	51	1.28	0.17	49	1.42	0.24	58	1.78
LM	3.43	12	1.28	0.67	9	1.42	0.73	8	1.78
factorization	0.0055	—	1.37	0.0026	—	1.97	0.0032	—	3.35

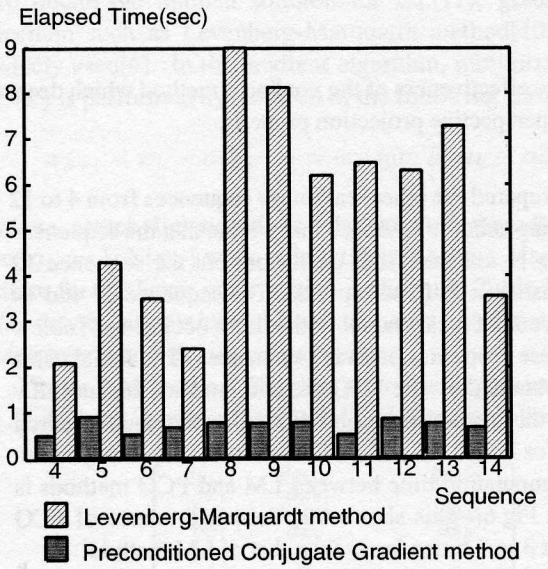


Figure 6: Comparison of computation times.

7 Conclusion

This paper describes an effective nonlinear optimization method to recover 3D shape and motion. To overcome the shortcomings, i.e. local minima and high computational cost, of the nonlinear approach, we propose "double search (DS) procedure" and "preconditioned conjugate gradient (PCG) algorithm". Experimental results using simulation and real data have demonstrated that the proposed method allows us to obtain the correct shape and motion with 3-9 times faster computation than the LM algorithm. Also we have shown the applicability of the method to a large building reconstruction from a set of partially tracked feature points.

In future, we will examine the performance of the method in various situations and also improve the algorithm to recover 3D shape and motion in realtime.

References

[1] H. C. Longuet-Higgins: "A computer algorithm for reconstructing a scene from two projections," Nature, 293:133-135, 1981.

[2] H. C. Longuet-Higgins: "Visual motion ambiguity," Vision Research 26(1):181-183, 1986.

[3] R. Szelisky and S. B. Kang: "Shape Ambiguities in Structure from Motion," IEEE Trans. PAMI, Vol 19, No. 5, 1997.

[4] C. Tomasi and T. Kanade: "Shape and motion from image streams under orthography: A factorization method," IJCV, 9-2, pa. 137–154, 1992.

[5] C.J. Poelman and T. Kanade: "A paraperspective factorization method for shape and motion recovery," IEEE Trans. PAMI, Vol. 19, pp. 206–218, 1997.

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

[7] E. V. E. Kumar, A. Tirumalai and C. Jain: "A Non-linear Optimization Algorithm for the Estimation of Structure and Motion Parameters," Proc. CVPR, pp. 136–153, 1989.

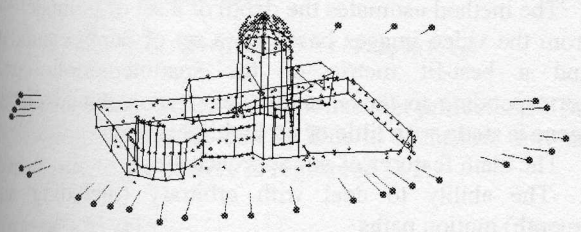
[8] J. Weng, N. Ahuja and T. S. Huang: "Optimal motion and structure estimation," IEEE Trans. PAMI, Vol. 15, No. 9, pp. 864–884, 1993.

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

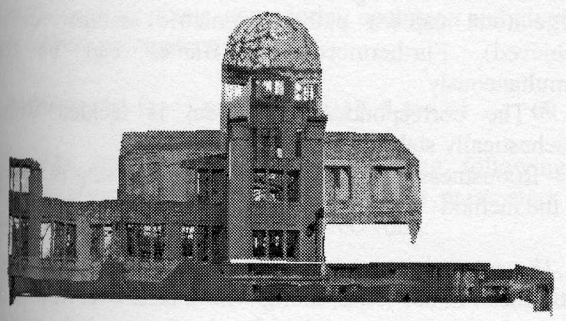
[10] K. Levenberg: "A Method for The Solution of Certain Non-linear Problems in Least Squares," *Quart. Appl. Math.*, vol. 2, pp. 164–168, 1944.

[11] D. W. Marquardt: "An Algorithm for Least-squares Estimation of Nonlinear Parameters," *SIAM J. Appl. Math.*, vol. 11, pp. 431–441, 1963.

(a) An example of real images in 29 frames

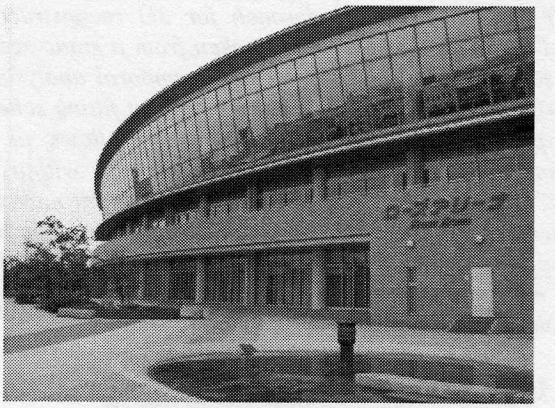


(c) Reconstructed shape with camera position (birds eye view)

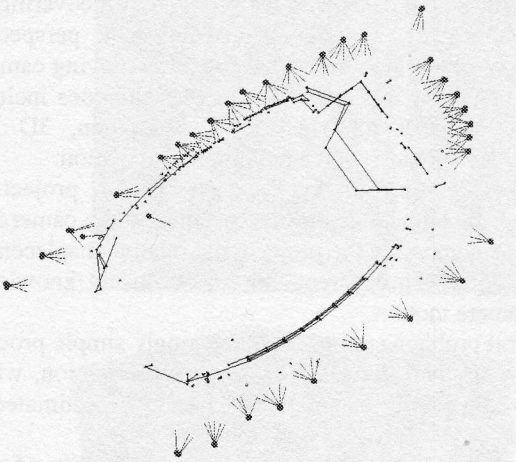


(d) Side view of the reconstructed dome with texture image.

Figure 7: The Hiroshima Atomic Bomb Dome.



(a) An example of real image in 144 frames



(b) Reconstructed shape with camera position (top view)

Figure 8: A stadium.