Paper

Fusion of 3D Shapes in Multiple View Points to Obtain More Accurate Shape

Non-member	Joon Bo Shim	(Nagoya University)
Non-member	Toshiharu Mukai	(RIKEN)
Member	Noboru Ohnishi	(Nagoya University, RIKEN)

The aim of this paper is to obtain an exact 3D model of a scene using multiple images from different camera positions. The reconstruction of 3D shape using two images has problem such as being weak at noise. Therefore we present methods to reduce noise and to improve the accuracy of 3D shape with multiple images. The system is divided into three stages: firstly, reconstruction of 3D shapes at different camera positions, secondly fusing the 3D shapes to obtain a para-ideal shape, and thirdly removing the outlier shapes and feature points by evaluation function, and fusing the rest of shapes. Even though the corruption of image data by noise is one of the unavoidable problems in any system, this paper shows how well the noise is removed by the proposed algorithms in multiple view points. We demonstrate a significant improvement of recovered shape. Experimental results show that our system performs well to remove noise with robustness. The maximum noise reduction rate is 82% in the real image experiment.

Keywords: Multiple View Points, Steepest Descent Method, Shape Fusion, Evaluation Function, Outlier

1. Introduction

In the field of computer vision, one of the main issues is that how exactly we can obtain the 3D shape of a target object from its perspective images. Many methods have been developed so far [1]~[4]. Those methods are useful to many fields such as virtual reality, shape recognition, navigation for a moving robot. And they are classified into stereo method [17] and shape-from-motion method [18]. We would like to focus on the latter method because this method does not need to give camera motion and is more suitable for application in daily life than the stereo method.

There are many methods to recover 3D shape and camera motion from a sequence of images [4]~[9] [13]~[15]. Among them, the factorization method shows good results [13]~[15]. This method, however, assumes the orthogonal or para-perspective projection and needs iterative calculation to find solutions. On the other hand, Mukai [8] proposed a linear method to estimate 3D motion and shape under perspective projection. Compared with the non-linear method, linear method has many merits: guaranteeing the uniqueness of solution, avoiding the iterative search and low computation load.

Therefore by using Mukai's method [8], we obtain 3D shapes which are represented in different coordinate systems conducted in multiple view points. These shapes are easy to be corrupted with noise. However as well known in statistics, integrating the shapes including noise can reduce the amplitude of noise in the shape [11]. Deguchi [15] also proposed improved factorization method under the perspective view in multiple view points. They used perspective projection to recover 3D shape without the approximation of orthogonal projection. It is said that known 3D points are provided to obtain the camera calibration values. Oliensis [12] dealt with noise in multi-frame structure. Motion error model is used to estimate 3D points error by the correlations between motion error and structure error.

In this paper, we propose methods to improve the accuracy of shape by fusing estimated 3D shapes and by the removal of outlier points and shapes in multiple view points. To obtain 3D motion and shapes, a camera moves around the object(the reverse case can be available). Then the method of fusion is conducted, which is based on that each obtained shape is the same shape and moving rigidly, however they are represented in a different coordinate system whose origin is moved in multiple view points. Therefore, we can match corresponding points, then convert each estimated shape to any standard coordinate system by coordinate transformation. Then we make the fusion of the transformed shapes to improve the accuracy of the object shape. We first present method for determining parameters of coordinate transformation based on feature points correspondences. Next, we show two methods to integrate shapes recovered in multiple view points. The result shape of integration is called a *para-ideal shape*, which is a refined shape. Then an evaluation function is applied to check out the outlier points and shapes by using the *para-ideal shape*. Outlier points in the image influence badly on the exact recovery of 3D shape. Therefore the elimination of outliers plays an important role to reconstruct 3D shape with high accuracy. The reconstructed inaccurate shapes are fused with the coordinate systems adjusted. The fusion of low noise leveled 3D shapes has

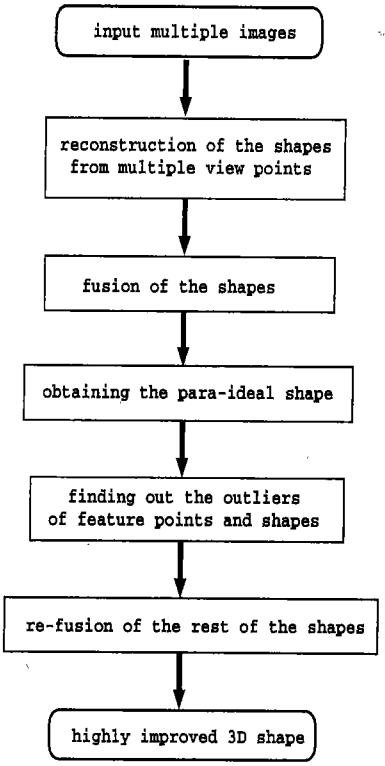


Fig. 1. The flow chart of algorithm.

much better advantages to reduce the noise than that of high noise leveled shape. In that sense, known the high noise leveled feature points and shapes called outliers by the evaluation function, we can obtain an exact shape by ruling out the outliers. Therefore the key points to obtain more exact shape is to find out the outliers and how to fuse the reconstructed 3D shapes. The noise of points in the real image has a property of point dependence. We mention three more methods to deal with outlier points and shapes and recovering shape by the elimination of outliers. Re-fusion of 3D shapes with less noisy level after the elimination of the outliers by the evaluation function shows excellent performance to reduce the noise in the real image experiment. Fig.1 shows the flowchart of the algorithm.

There are a few conventional statistical fitting methods reported such as M-estimators method and LMedS [20][21] in computer vision. The different thing between our method and conventional methods is to adjust the coordinate system to fit the shapes. Therefore a *para-ideal shape* is proposed as a standard coordinate system and a criterion of evaluation function.

We explain Mukai[8]'s algorithm to reconstruct 3D shape in section 2. Then we present two methods of shapes fusion in section 3. Section 4 is about the removal of the outliers by the evaluation function and re-fusion of 3D shapes. And we describe the simulation and experiment in section 5 and 6. We finally conclude this paper.

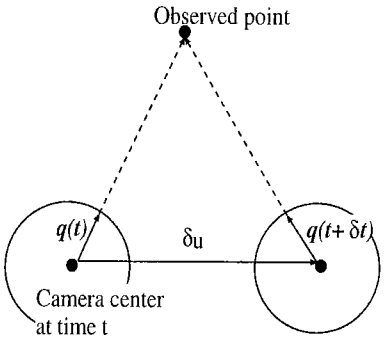


Fig. 2. Relationship of camera positions before and after infinitesimal time lapse.

2. Obtaining 3D Shape [8]

2.1 Spherical Imaging Model In order to simplify the computation and obtain clear perspective, a spherical image screen is used in the system. It is reported that self-motion is more easily estimated in the case of a spherical image screen than in the case of a planar screen [16]. A point $\boldsymbol{r}$ is projected on the spherical screen with unit radius $\boldsymbol{r}/||\boldsymbol{r}||$.

We set two coordinate systems : a camera coordinate system and a world coordinate system. Let $\boldsymbol{q}$ and $\tilde{\boldsymbol{q}}$ be vectors representing the projected points $(\boldsymbol{r}/||\boldsymbol{r}||)$ in the world and camera coordinate system, respectively. Without losing generality, let the world coordinate system be set to coincide with the camera coordinate system just before the camera moves. Then, $\boldsymbol{q}$ and $\tilde{\boldsymbol{q}}$ are expressed as

$$\boldsymbol{q} = \tilde{\boldsymbol{q}} = \frac{\boldsymbol{r}}{||\boldsymbol{r}||} \quad \cdots\cdots (1)$$

Though $\boldsymbol{q}$ and $\tilde{\boldsymbol{q}}$ are the same, their time derivatives $\dot{\boldsymbol{q}}$ and $\dot{\tilde{\boldsymbol{q}}}$ are different from each other owing to the camera's motion. We also have the equation

$$\dot{\boldsymbol{q}} = \dot{\tilde{\boldsymbol{q}}} + \boldsymbol{\omega} \times \tilde{\boldsymbol{q}}, \quad \cdots\cdots (2)$$

where $\times$ denotes the outer product and $\boldsymbol{\omega}$ is the angular velocity of the camera with respect to the world coordinate system. On the other hand, the accurate reconstruction of 3D shape on the perspective image is up to obtaining accurate optical flow($\dot{\tilde{\boldsymbol{q}}}$). However the difficulty in obtaining accurate optical flow is a major problem in computer vision. We assume that optical flow of feature points is already obtained and express how motion and structure are recovered from the optical flow in the next section.

2.2 Recovery of Shape with Motion In order to obtain the position of feature points, we must find the camera motion, translation vector $\boldsymbol{v}$ and angular velocity $\boldsymbol{\omega}$. Fig.2 shows the camera movement $\delta\boldsymbol{u}$ in infinitesimal time lapse δt. $\boldsymbol{q}(t)$ and $\boldsymbol{q}(t+\delta t)$ are projective points at time t and $t+\delta t$. Because $\boldsymbol{q}(t)$, $\boldsymbol{q}(t+\delta t)$ and $\delta\boldsymbol{u}$ are in the identical plane, we can obtain Eq.(3).

$$((\dot{\tilde{\boldsymbol{q}}} + \boldsymbol{\omega} \times \tilde{\boldsymbol{q}}) \times \tilde{\boldsymbol{q}}) \cdot \boldsymbol{v} = 0. \quad \cdots\cdots (3)$$

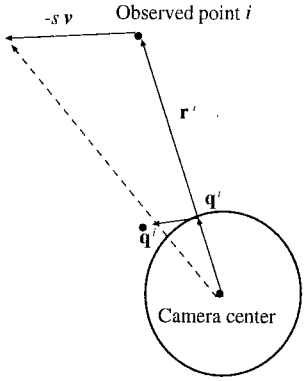


Fig. 3. Geometry of an observed point and the camera.

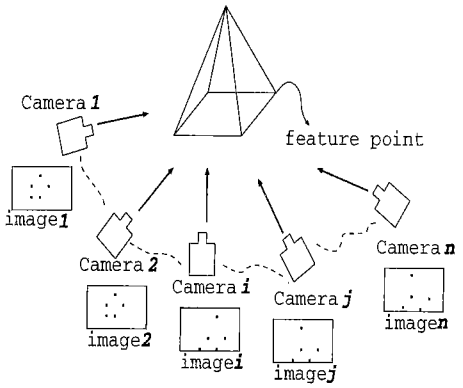


Fig. 4. Image sequence obtained from different view points by a moving camera.

When $\boldsymbol{v} \neq 0$ ($\boldsymbol{v} = 0$ is not dealt with), by expanding Eq.(3), we obtain

$$(\dot{\tilde{\boldsymbol{q}}} \times \tilde{\boldsymbol{q}}) \cdot \boldsymbol{v} - \boldsymbol{\omega} \cdot \boldsymbol{v} + (\tilde{\boldsymbol{q}} \cdot \boldsymbol{\omega})(\tilde{\boldsymbol{q}} \cdot \boldsymbol{v}) = 0. \quad \cdots\cdots \quad (4)$$

We can provide a better approach to solve Eq.(4) by setting $\boldsymbol{b} \equiv \dot{\tilde{\boldsymbol{q}}} \times \tilde{\boldsymbol{q}}$ and $\boldsymbol{A} \equiv \tilde{\boldsymbol{q}}\tilde{\boldsymbol{q}}^T - I$, where I is the 3×3 identity matrix. We can simplify the Eq.(4) as

$$\boldsymbol{b}^T\boldsymbol{v} + \boldsymbol{\omega}^T\boldsymbol{A}\boldsymbol{v} = 0. \quad \cdots\cdots\cdots\cdots\cdots\cdots \quad (5)$$

Using the symmetry of the matrix $\boldsymbol{A}$, Eq.(5) can be rewritten as

$$[\boldsymbol{b}^T \boldsymbol{d}^T] \begin{bmatrix} \boldsymbol{v} \\ \boldsymbol{m} \end{bmatrix} = 0, \quad \cdots\cdots\cdots\cdots\cdots\cdots \quad (6)$$

where $\boldsymbol{d} = [a_{11}\ a_{22}\ a_{33}\ a_{12}\ a_{23}\ a_{31}]^T$, a_{ij} is a component of the matrix $\boldsymbol{A}$ and $\boldsymbol{m}$ is expressed as

$$\boldsymbol{m} = \begin{bmatrix} \omega_1 v_1 \\ \omega_2 v_2 \\ \omega_3 v_3 \\ \omega_1 v_2 + \omega_2 v_1 \\ \omega_2 v_3 + \omega_3 v_2 \\ \omega_3 v_1 + \omega_1 v_3 \end{bmatrix}. \quad \cdots\cdots\cdots\cdots\cdots \quad (7)$$

Because the dimension of unknown vectors $\boldsymbol{v}$ and $\boldsymbol{m}$ is nine, we need at least eight points to obtain the unknowns. If n points are used, $\boldsymbol{v}$ and $\boldsymbol{m}$ are determined by calculating the null space of a $n\times9$ matrix. Then $\boldsymbol{\omega}$ is computed by substituting the obtained $\boldsymbol{v}$ and $\boldsymbol{m}$ into Eq.(7). Once $\boldsymbol{v}$ and $\boldsymbol{\omega}$ are determined, $\boldsymbol{r}$ can be calculated by using the law of similarity in triangle(see Fig.3).

$$\boldsymbol{r}^i = -s\frac{\boldsymbol{v} \cdot \dot{\boldsymbol{q}}^i}{||\dot{\boldsymbol{q}}^i||^2}\boldsymbol{q}^i, \quad (i = 1, 2, \cdots, n), \quad \cdots\cdots\cdots \quad (8)$$

where s is a scale factor inevitable in the problem of shape-from-motion.

3. Fusion of Obtained Shapes

3.1 Coordinate Transform We can obtain shapes $\boldsymbol{r}_j$ from different view points by camera movement as shown in Fig.4. In order to improve the accuracy of obtained shape, we propose how to integrate the obtained shapes at each of the different view points. When a camera moves, a new coordinate system is made at each view point. Each obtained shape is represented by the different coordinate systems. This means that each coordinate system of the recovered shape has a different scale, rotation and translation. Therefore the integration of obtained shapes is not simple. Hence, we need to convert the shape $\boldsymbol{r}_j$ in the j-th coordinate system into the shape $\boldsymbol{r}_{j\to standard}$ in a standard coordinate system as follows:

$$\boldsymbol{r}_{j\to standard} = s\boldsymbol{R}\boldsymbol{r}_j + \boldsymbol{t}, \quad \cdots\cdots\cdots\cdots\cdots \quad (9)$$

where s, $\boldsymbol{R}$, $\boldsymbol{t}$ mean scale, rotation matrix and translation vector, respectively. Rotation matrix $\boldsymbol{R}$ is expressed below. To obtain rotation matrix $\boldsymbol{R}$, each angle α, β and γ around the $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$ axes should be calculated.

$$\begin{aligned} \boldsymbol{R} &= \boldsymbol{R}(\boldsymbol{z}, \gamma)\boldsymbol{R}(\boldsymbol{y}, \beta)\boldsymbol{R}(\boldsymbol{x}, \alpha) \\ &= \begin{bmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix} \\ &\quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{bmatrix} \end{aligned}$$

Given the correspondence of at least eight feature points between both coordinate systems, we can find the parameters s, $\boldsymbol{R}$ and $\boldsymbol{t}$ by minimizing the following criterion:

$$E(s, \boldsymbol{R}, \boldsymbol{t}) = \sum_{i=1}^{n} ||\boldsymbol{r}^i_{standard} - \boldsymbol{r}^i_{j\to standard}||^2, \quad (10)$$

where i denotes i-th feature point of an object, and n is the number of feature points. We used the steepest descent method to minimize the criterion, Eq.(10).

3.2 Fixed-Frame Fusion Method By using Eqs.(9) and (10), all the shapes are converted to the standard coordinate system, which means that $\boldsymbol{r}^i{}_{standard}$ is replaced by $\boldsymbol{r}^i{}_1$ into Eq.(10). Then we can obtain Eq.(11).

$$r_{1\to standard}, r_{2\to standard}, \cdots, r_{j\to standard} \quad \cdots (11)$$

, where $r_{j\to standard}$ means a converted shape of r_j into the standard coordinate system as $r_{standard}$(the candidate of standard coordinate system can be set to any shape such as $standard$=1,2,$\cdots$,j). If the coordinate system of r_1 is selected as a standard coordinate system, we integrate r_1 and the converted ones $r_{2\to1}\cdots r_{j\to1}$ as follows :

$$f_j = \frac{r_1 + r_{2\to1} + \cdots + r_{j\to1}}{j}, \quad \cdots\cdots\cdots\cdots (12)$$

where $j = 1, 2, \cdots$, $end\ of\ shape$. This equation is rewritten into a recursive form:

$$f_j = \frac{(j-1)\cdot f_{j-1} + r_{j\to1}}{j}, \quad \cdots\cdots\cdots\cdots (13)$$

where f_0=0.

3.3 **Renewing Frame Fusion Method** This method fuses shapes one by one by integrating the j-th shape with the latest fused shape f_{j-1} as shown in Eq.(14).

$$f_j = \frac{(j-1)\cdot f_{j-1} + r_{j\to f_{j-1}}}{j}, \quad \cdots\cdots\cdots\cdots (14)$$

where $j = 1, 2, \cdots$, $end\ of\ shape$ and $r_{j\to f_{j-1}}$ is a converted coordinate by corresponding the feature points between r_j and the fused shape f_{j-1}, which means ${r^i}_{standard}$ is replaced by ${f^i}_{j-1}$ into Eq.(10). Note that $r_{1\to0}$=r_1. The difference between Eq.(13) and Eq.(14) is the second term. Comparing them, $r_{j\to f_{j-1}}$ is expected to be more accurate than $r_{j\to1}$ because the shape f_{j-1} is the result of integrating $(j-1)$ shapes and contains lower noise than the shape r_1 of standard coordinate system by Fixed-frame fusion method.

4. Outlier Points/Shapes Removal and Re-fusion

Some of outliers corrupt entire recovered shape. In this section, we explain how we can find the outliers efficiently. Then, re-fusion process is explained in the last section. There are two parts to check the outliers out from each shape. Firstly, we deal with the feature points. Secondly, noisy shapes are checked out as the outlier shapes.

Note that we apply the process of removal of outliers to only real image experiment since we add the uniformly random noise to each point in simulation. It means the outliers in simulation are point independent and isotropic Hence, each point has isotropic noise at each movement of camera, we cannot expect the satisfactory results in simulation. However, in the real image, the noise distribution is heteroscedastic(point dependent), and anisotropic. The location of feature point influences on the magnitude of noise in each point. It is of no use including outliers in fusion to improve accuracy.

4.1 **The Points-Based Selection of Better Feature Points** Any noisy feature points are much farther away from their corresponding feature points of ideal shape than other less noisy feature points. As we addressed in previous chapter, we could improve the accuracy of shape by the fusion of shapes. Therefore we assume that the last refined fused shape f_j^i is regarded as an ideal shape, where j means the number of shapes. The distance between each feature point (r_j^i) and para-ideal feature points(f_j^i) is a kind of criterion to rule out the outliers. We define the evaluation function to select noisy points as

$$\Phi_j^i = (r_j^i - f_j^i)^T (V_j)^{-1} (r_j^i - f_j^i), \quad \cdots\cdots\cdots (15)$$

where i is the point number, j is the index of each shape and V_j is the covariance matrix of the error $e_j^i = r_j^i - f_j^i$.

Then we obtain error of each feature point by Eq.(15). If any points has a higher error than the threshold, they are outliers which should be ruled out. Selected outliers have about double times more over error than avarage error. The criterion and cut off line may be changed according to the system's characteristic such as two or three times more over than average error. We show the heteroscedastic nature of feature points in experiment of later section.

4.2 **Removal of Outlier Shapes** To find the outlier shapes, we calculate their Euclidean distance between each shape and a para-ideal shape(f_j) by Eq.(16).

$$Error\ by\ Para - Ideal = \sum_{i=1}^{n} ||f_j^i - r_j^i||/n, \quad (16)$$

where ${r^i}_j$ is the each recovered shape. j means the number of shapes. n is the number of feature points. We set the threshold which is 1.19 times more over than average error. Eq.(16) gives us the error level of each shapes shown in Fig.(13). If any shape's error is 1.19 times more over than average error, the shape is regarded as an outlier shape to be ruled out in this paper. The criterion and threshold may be changed according to the system's characteristic as the points-based outlier selection.

4.3 **Re-fusion of the Rest of the Shapes** Relatively higher noise leveled shapes and feature points are selected as outliers. Outliers are the targets to be ruled out at the step of re-integration of the shapes over again. It is worthy of notice that where we remove the outlier in the image, or in the shape after the recovery of shape. As for the noise reduction efficiency, removal of outlier points in the image is better than in the shape because the outliers corrupt others in the 3D reconstruction. The re-integration of the shapes is applied by Renewing frame fusion method mentioned in section 3.

5. Simulation

5.1 Condition

(1) As shown in Fig.5, 214 feature points are located in the surface of a cube($20 \times 20 \times 20$), and

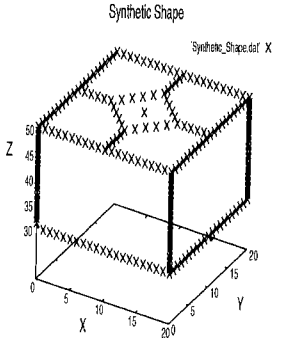


Fig. 5. Synthetic Shape.

30-50 units apart from the camera center along the z-axis. Visual angle is not specified. The scale cannot be specified because of no existence of scale.

(2) Camera moves around the object on the circumference of a circle, the radius of which is the distance between the center of object and the origin of camera coordinate. Then rotation $\boldsymbol{\omega}$ is [0, $\frac{\pi}{180}$rad, 0] and translation $\boldsymbol{v}$ is the distance of one degree moving of a camera.

(3) We add uniform random noise to the optical flow of points after projection on the screen. Noise amplitude is less than 1% in comparison to the optical flow. Each point is perturbed as follows:

$$\dot{\tilde{\boldsymbol{q}}}_{noised} = \dot{\tilde{\boldsymbol{q}}}_{pure} + \boldsymbol{n}.$$

5.2 Method

(1) We conducted simulation with two methods: Fixed-frame fusion method and Renewing frame fusion method.

(2) We calculated the errors per a point.

$$Error = \sum_{i=1}^{n} ||\boldsymbol{r}^i{}_{ideal} - \boldsymbol{f}^i{}_j||/n, \quad \cdots\cdots \quad (17)$$

where $\boldsymbol{r}^i{}_{ideal}$ is the ideal 3D position of each feature points, and $\boldsymbol{f}^i{}_j$ is the obtained one after fusion by Eqs.(13) or (14). j means the number of fused shapes which is 12 in the present paper. n is the number of feature points.

(3) We calculated the rate of error reduction.

$$Rate = \frac{(\boldsymbol{E}_{\boldsymbol{r}_1} - \boldsymbol{E}_{\boldsymbol{f}_{latest}})}{\boldsymbol{E}_{\boldsymbol{r}_1}} \times 100(\%), \quad \cdot \quad (18)$$

where $\boldsymbol{E}_{\boldsymbol{r}_1}$ is the error per point of the shape obtained firstly and $\boldsymbol{E}_{\boldsymbol{f}_{latest}}$ is the error per point of the latest fused shape.

5.3 Results and Discussion Figs.6~8 show the results. In Fig.6, horizontal axis shows the number of fusion and vertical axis shows the error per point defined in Eq.(18). For 'Each Shape' in the figure, the horizontal axis is each camera position. Then the plot shows the error of the shape recovered at each view point. Fig.8 is the reconstructed shapes and fused shapes. Seeing the plot of 'Each Shape' in Figs.6 and 8, we find that the error of each shape unsteadily vibrates depending on camera position. By fusing shapes at the different view points, we find that error decreases lower than half of those before the shape integration. The rate of error reduction is 58.5% by Renewing frame fusion method.

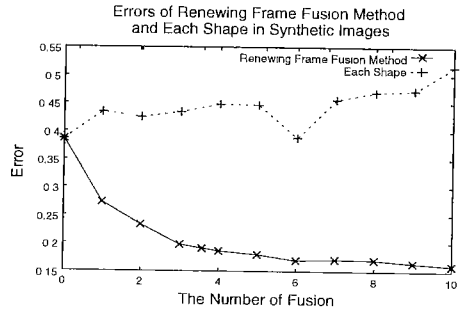


Fig. 6. Error plots by Renewing frame fusion method in synthetic image.

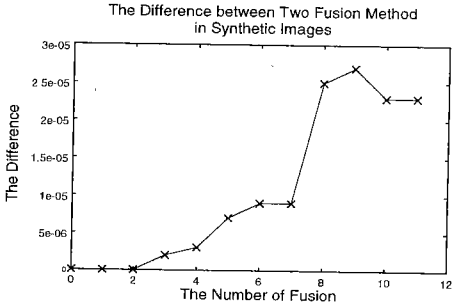


Fig. 7. The difference between the error of two fusion methods which are Fixed-frame fusion method and Renewing frame fusion method in synthetic image. The difference is obtained by subtracting the error of Fixed-frame fusion method from that of Renewing frame fusion method.

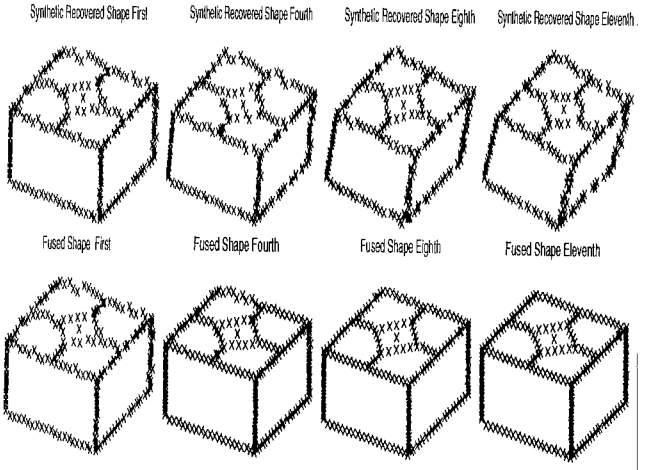


Fig. 8. Reconstructed shapes(the upper row) and fused shapes(the lower) by Renewing frame fusion method in synthetic images.

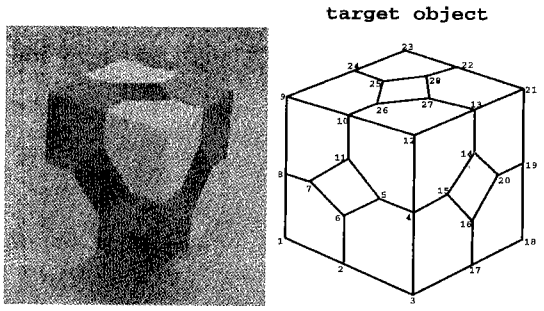


Fig. 9. An object used in the real image experiment.

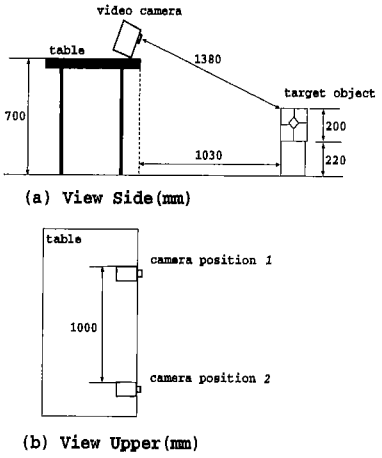


Fig. 10. Experimental Environment.

Comparing two fusion methods mentioned in section 3, there is a slight difference between two methods. Fig.7 is the error difference between the proposed two fusion methods. As the number of fusion increases, the difference between two fusion methods is increased. Its difference is obtained by subtracting the error of Fixed-frame fusion method from Renewing frame fusion method. However the difference is too small to confirm its validity in simulation.

In simulation, we did not apply the method of removal of outliers since we add uniformly random noise to each points. By addition of uniformly random noise to feature points, each shape has different random outliers, therefore it is hard to determine the correspondence of feature points between two images or points to integrate shapes. It means the outliers are point independent and isotropic. Therefore, the isotropic nature of noise makes equivalent noise distribution at each movement of camera. We cannot expect the satisfactory results of outlier removal process in simulation. However, according to the fusion of shapes without process of removal of outliers, each fused shape gradually lessen its noise as shown in Fig.6.

6. Experiments

6.1 Condition

(1) Experimental situation is shown in Fig.10. Actually a camera(Sony PC-100) takes an object with free motion. However in this experiment, to guarantee its enough translation and no occlusion, a camera moves around from position 1 to position 2 repeatedly. Camera rotation is set to be different randomly in each camera's movement.

(2) The number of feature points: 28 (Fig.9.)

(3) The number of used images: 21 (The number of reconstructed shapes is 20.)

(4) Camera parameters used to convert the planar image into spherical image : Tsai [19]'s method is used.

6.2 Methods

(1) According to applying the process of removal of outliers, we divide into five methods.

- Method 1: Fixed-frame fusion method only without the process of removal of outliers.
- Method 2: Renewing frame fusion method only without the process of removal of outliers.
- Method 3: Outliers are not removed in the reconstructing process, but in the process of the fusion. Then shapes are fused by Renewing frame fusion method.
- Method 4: Outlier points are removed from the image data, then shapes are reconstructed. Then shapes are fused by Renewing frame fusion method.
- Method 5: Outlier points are removed from image data, also outlier shapes are removed, then, Renewing frame fusion method is used.

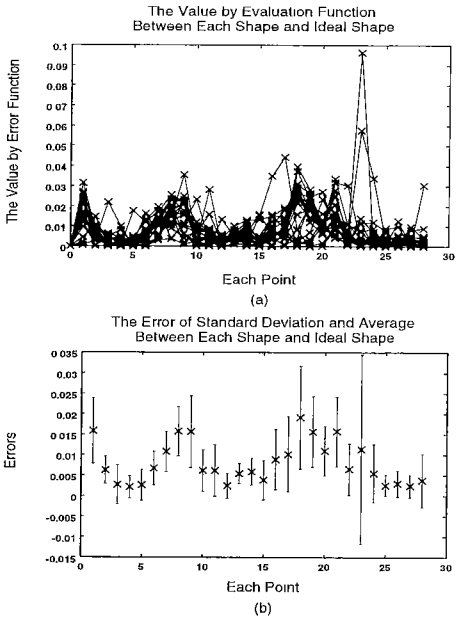


Fig. 11. Error plots by Eq.(15). x-axis is the order of points' number and y-axis is the value of evaluation function Eq.(15) between each shape and ideal shape in the real image experiment.

6.3 Results

6.3.1 The Removal of Outliers In Figs.11 and 12, (*a*) shows the errors of each feature point calculated

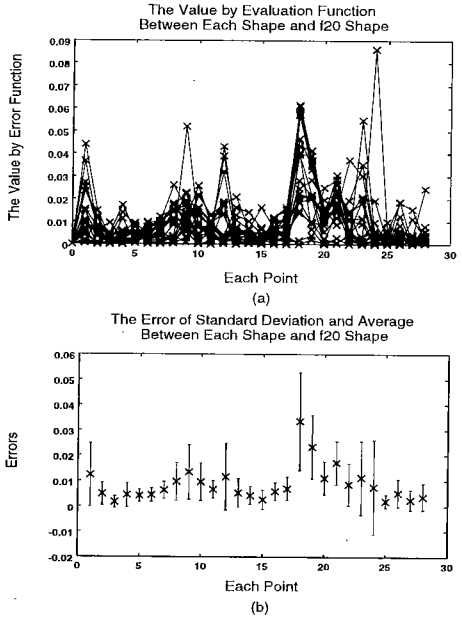


Fig. 12. Error plots by Eq.(15). x-axis is the order of points' number and y-axis is the value of evaluation function Eq.(15) between each shape and fused shape, f_{20} in the real image experiment.

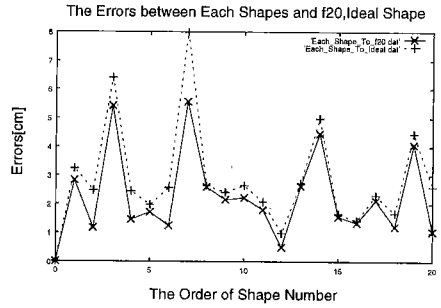


Fig. 13. Error plots of each shape to f_{20} and ideal shape. x-axis is the order of shapes' number and y-axis is the error per one point.

by Eq.(15) and ($\boldsymbol{b}$) is the average and the standards deviation of errors. The error in Fig.11 is between each shape($\boldsymbol{r}_j$) and an ideal shape, and the error in Fig.12 is between each shape and a para-ideal shape($\boldsymbol{f}_{20}$). There are similarities between two plots as shown in Figs.11 and 12. Hence, we can have a firm belief to select the outliers by using $\boldsymbol{f}_{20}$ instead of an unknown ideal shape. We regarded points 18, 19 and 21 as outliers by threshold(double times more over than average error).

Next step is to remove the outlier shapes. Fig.13 shows errors between each shape and ideal(or para-ideal) shape calculated by Eqs.(16) and (17). Some of shapes include much noise than others. We ruled out the five shapes($\boldsymbol{r}_1$,$\boldsymbol{r}_3$,$\boldsymbol{r}_7$,$\boldsymbol{r}_{14}$,$\boldsymbol{r}_{19}$) in the twenty shapes. All of their errors are 1.19 times more over than the average error(2.34cm). Therefore 1.19 times the average error can be used as threshold to obtain outlier shapes.

Seeing Fig.13, it is said that the errors of each shape have almost the same tendency between $\boldsymbol{f}_{20}$ and ideal shape. Therefore fused shape $\boldsymbol{f}_{20}$ can be used as the

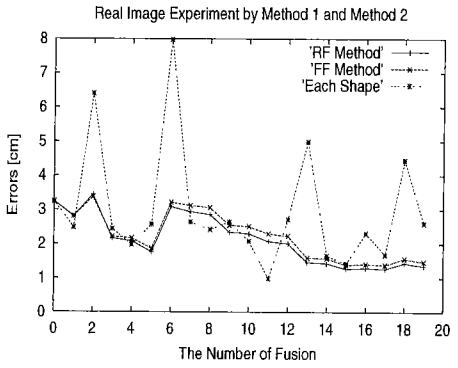


Fig. 14. The result of real images by methods 1 and 2. 'RF Method' means Renewing frame fusion method. 'FF Method' is Fixed frame fusion method. 'Each Shape' means the error of each shape.

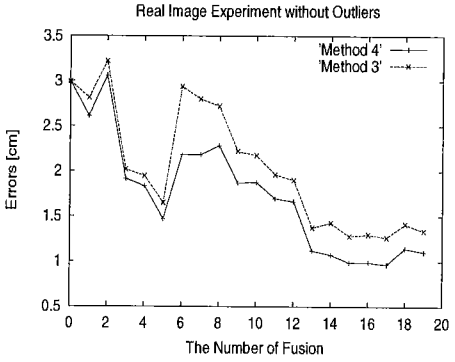


Fig. 15. The result of real images by removing outlier feature points. Outlier feature points are removed in the fusion process for method 3, while in the reconstruction process for method 4.

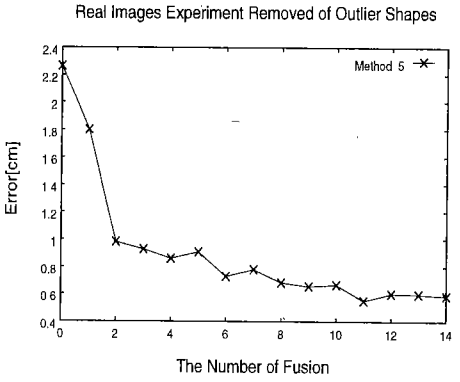


Fig. 16. The result of real images without outlier shapes by method 5. The number of fusion is reduced by five comparing to Figs.15 and 16 since five outlier shapes are removed in fusing shapes.

criterion shape to remove the outlier shapes.

6.3.2 Error Comparison among the Methods

We conducted the experiments of five methods concerning about removal of outlier feature points explained in section 4 and section 6.2. Fig.14 shows the point er-

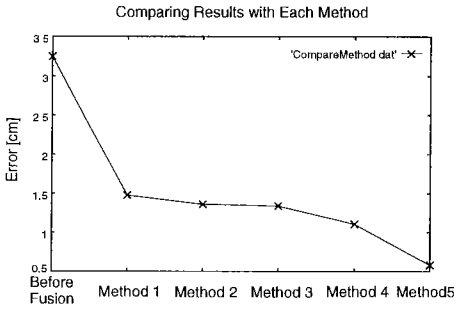


Fig. 17. Comparison of final error among the methods.

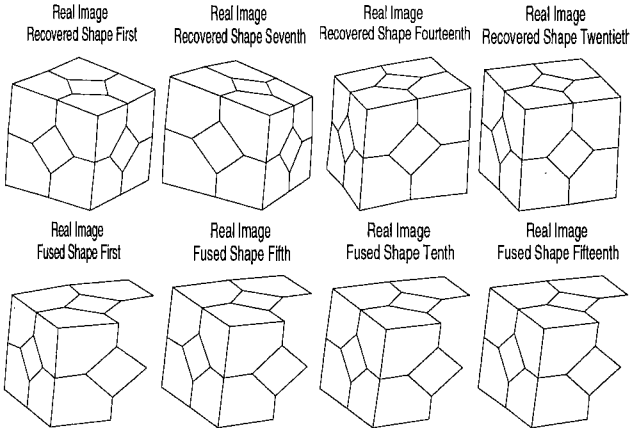


Fig. 18. Upper four shapes: reconstructed shapes without fusion method at each view point. Bottom four shapes: fused shapes by method 5. There are lost three points in the shapes as outliers.

rors of method 1, 2 and Each shape. The rates of error reduction are 54.4%(method 1), 58.1%(method 2).

Fig.15 shows the result of method 3 and 4. The rates of error reduction are 58.8%(method 3), 66.0%(method 4). Thus the method 4 is better than the method 3. The reason is that the outlier feature points corrupt others when a 3D shape is reconstructed. Hence, in the step of obtaining optical flow, the removal of outlier points is helpful to obtain better exact 3D shape.

The method 5 removes both outlier shapes and feature points. $\boldsymbol{f}_{20}$ is used as a para-ideal shape. Fig.16 shows its result. The shape number,1,3,7,14 and 19 are ruled out because they have over the threshold error which is 1.19 times more than average error. The outlier feature points also are ruled out in reconstructing 3D shape from image data. The rate of error reduction is 82%.

As shown in Fig.17, the rates of error reduction are 54.4%, 58.1%, 58.8%, 66.0% and 82.0%, which are Fixed-frame fusion method(method 1), Renewing frame fusion method(method 2), method 3, method 4, method 5, respectively. The method 5 is the best performance in error reduction.

6.3.3 **Fused Shapes** Fig.18 shows the reconstructed shapes and fused shapes. Upper four shapes are each reconstructed shapes at different view points. Lower four shapes are corresponding fused shapes. We find that each reconstructed shapes are improved by shape fusion.

6.4 **Discussion** In the real image experiment, the noise distribution is heteroscedastic(point dependent), and anisotropic as shown in Figs.11, 12 and 14. The location of feature point influences on the magnitude of noise in each point. As shown in Figs.11 and 12, feature point number 1, 8, 9, 18, 19, 21 (see Fig.9) are easy to include noise in the experiment. It seems that the vicinity of the edge is prone to contain the error of optical flow.

Therefore the feature point number 18, 19 and 21 are removed as outliers by the process of removal of outliers. The three points have more than double times error comparing to the average error by evaluation function in Eq.(15).

In the reconstructing shapes, outlier feature points perturb stable recovery of other points. Therefore, in the step of obtaining 3D shape, removal of outliers plays an important role on exact recovery of 3D shape.

7. Conclusion

In this paper, we have proposed the methods to improve the accuracy of obtained shape focusing on outliers, and investigated the rate of error reduction by these methods. By fusing obtained shapes and evaluation function, we could decrease the position error by 58.5% in the simulation and 82.0% in the experiment. The fusion can smooth noise and give us more stable shape although the error of each recovered shape fluctuates depending on each viewing point. As shown in Fig.18, there are lost three points as outliers. However, another image sequence obtained from new view points may make it possible to obtain the lost three points with high accuracy. Therefore lost three points can be supplied from another set of fused shapes. Finding a integration method to supply lost points and another better methods of fusion to reduce noise are future subjects.

Acknowledgment One of authors Joon Bo Shim, expresses his thanks to Hori Information Science Promotion Foundation for supporting his research.

(Manuscript received March 24, 2000, revised July 7, 2000)

References

(1) Zhang,Z., "Estimating motion and structure from correspondences of line segments between two projective images," IEEE Trans. Pattern Anal. Machine Intell., Vol.PAMI-17, No.12, pp.1129-1139 (1995).

(2) Zhang, Z. and Xu,G., "A unified theory of uncalibrated stereo for both perspective and affine camera," International of Journal of Computer Vision, Vol.9, pp. 213-229 (1998).

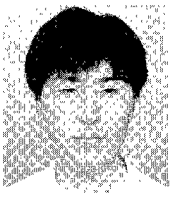
(3) Sato, J. and Cipolla, R., "Uncalibrated reconstruction of curved surface," Image and Vision Computing 17 pp.617-623 (1999).

(4) Tsai, R. and Huang, T., "Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces," IEEE Trans. Pattern Anal. Machine Intell., Vol.PAMI-6, No.1, pp.13-27 (1984).

(5) Tagawa, N. and Totiu, T. and Endoh, T., "Estimation of 3-D motion from optical flow with unbiased objected function," IEICE Trans Inf.& Syst., Vol.E77-D, No.10, pp.1148-1161 (1994).
(6) Wang, S. and Huang, C., "Structure recovery and motion estimation from stereo motion," IEICE Trans Inf.& Syst., Vol.E77-D, No.11, pp.1247-1258 (1994).
(7) Kanatani, K. and Takeda, S., "3-D motion analysis of a planar surface by renormalization," IEICE Trans Inf. & Syst.,Vol.E78-D, No.8, pp. 1074-1079 (1995).
(8) Mukai, T. and Ohnishi, N., "Motion and shape from perspectively projected optical flow by solving linear simultaneous equations", Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems(IROS'97), pp.740-745, Grenoble, Sep 7-11 (1997).
(9) Cipolla, R. and Blake, A., "Surface shape from the deformation of apparent contours," International Journal of Computer Vision, Vol.no.9, Issue No.2, pp.83-112 (1992).
(10) Horn, B., "Robot vision," The MIT Press, (1986).
(11) Clark, J. and Yuille, A., "Data fusion for sensory information processing system," Kluwer Academic Publisher (1990).
(12) Thomas, J. and Oliensis, J., "Dealing with Noise in Multi-frame Structure from Motion," Computer Vision and Image Understanding, Vol.76, No.2,November, pp.109-124 (1999)
(13) Tomasi, C. and Kanade, T., "Shape and motion from image streams under orthography: A factorization method," International of Journal of Computer Vision, Vol.9, pp. 137-154 (1992).
(14) Poelman, C. and Kanade, T., "A paraperspective factorization method for shape and motion recovery," IEEE Trans. Pattern Anal. Machine Intell., Vol.PAMI-19, No.3, pp.206-218 (1997).
(15) Deguchi, K., "Factorization method for structure from perspective multi-view images," Sensing Instrument Control Engineering Systems Information Computer Ergonomics, Vol.34, No.10, pp.1321-1328 (1998).
(16) Fermuller, C. and Aloimonos, Y., "Ambiguity in structure from motion: Sphere versus plane," International Journal of Computer Vision, Vol.No.28, Issue No.2, pp.137-154 (1998).
(17) Bascle, B. and Deriche, R., "Stereo matching, reconstruction and refinement of 3D curves using deformable contours," Proceedings of ICCP, pp.421-430, Berlin (1993).
(18) Ohta, N., "Optimal structure-from-motion algorithm for optical flow," IEICE Trans.,Vol.E78-D,No.12, pp.1129-1139 (1995).
(19) Roger, 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. RA-3, No.4, pp.323-344 (1987).
(20) Rousseeuw, P. J., "Least Median of Squares Regression," Journal of American Statistical Association, Vol.79, pp.871-880 (1984)
(21) Rousseeuw, P. J. and Leroy, A. M., "Robust Regression and Outlier Detection," John Wiley & Sons (1986)

Joon Bo Shim (Non-member) He received the B.Eng. in electrical engineering from Kookmin University, Seoul, Korea in 1995. He had worked at Eastman Kodak Korea in 1995. He joined the information engineering in Nagoya University as a research student in 1996. He received the M.Eng. in the information engineering from Nagoya University in 1999. He is currently working for Ph.D degree at the same place. His research interests cover computer vision and image processing. He received Promotion Award from the 1999 Tokai-section joint conference of the seven institutes of electrical and related engineers in 2000. He is a member of the institute of electronics information and communication engineers, the institute of image information and television engineers, the robotics society of Japan

Toshiharu Mukai (Non-member) He received the B.Eng., the M.Eng. and the Dr.Eng. degrees in mathematical engineering and information physics from the University of Tokyo in 1990, 1992 and 1995, respectively. He has been a frontier research scientist at the Bio-Mimetic Control Research Center under the Institute of Physical and Chemical Research (RIKEN), Japan, since 1995. His current research interests include sensor fusion, active sensing, shape recovery from an image sequence, and neural network. He got the 1999 Best Paper Award from Society of Instrument and Control Engineers (SICE) and the 2000 Award for Research Achievement from the Virtual Reality Society of Japan.



Noboru Ohnishi (Member) He received the B. Eng., M. Eng. and D. Eng. degrees from Nagoya University, Nagoya, Japan, in 1973, 1975 and 1984, respectively. From 1975 to 1986 he was with the Rehabilitation Engineering Center under the Ministry of Labor. From 1986 to 1989 he was an Assistant Professor in the Department of Electrical Engineering, Nagoya University. From 1989 to 1994, he was an Associate Professor. He is a Professor in the Department of Information Engineering, and concurrently a Head of Laboratory for Bio-mimetic Sensory System at the Bio-mimetic Control Research Center of RIKEN. His research interests include computer-vision and -audition, robotics, bio-cybernetics, and rehabilitation engineering. Dr. Ohnishi is a member of IEEE, IEEJ, IEICE, IPSJ, SICE, JNNS, IIITE and RSJ.