Space Time Carving for Shape and Motion Recovery

Fumihiko Sakaue, Masaaki Takami and Jun Sato

Nagoya Institute of Technology

Abstract—**Recently, a lot of methods for shape and motion reconstruction have been studied extensively. This paper describes a method for simultaneous shape and motion reconstruction without any corresponding points. We call the method space-time carving. The proposed method is based on a space carving method which can reconstruct 3D shape without any corresponding points. The method uses photo-consistency for 3D space carving. Our proposed method also uses photo-consistency for carving high dimensional space which represents not only shape but also motion. As a result, the proposed method can recover not only 3D shapes but also 3D motions of objects. The experimental results show that the proposed method works well.**

Index Terms—**Space carving, extended multiple view geometry, motion analysis**

I.   INTRODUCTION

The recovery of object shape and motion is one of the most important problems on computer vision. Thus, a lot of methods are studied in this field [1，2]. Recently, shape and motion recovery can be simultaneously accomplished by using extended multiple view geometry[3]. In this framework, an *N*-dimensional vector which includes shape and motion information is recovered from camera images, where $N \geq 3$. Thus, shape and motion are recovered simultaneously as in the case of shape recovery.

However, most of the methods for recovery of shape/motion need corresponding points in camera images. Although it is not so hard to find corresponding points for sparse object shape, it becomes a very hard problem when we would like to recover dense shape. In order to avoid the corresponding problem, Kutulakos el al.[6] proposed space carving method for 3D shape recovery. This method pays attention to ``photo consistency'' on object surface, and does not require explicit corresponding points on input images. However, the method cannot recover moving objects because the method assumes that input scenes are static. Although some methods which can recover moving object based on concept of space carving [4，5], the methods do not use motion information, directly for shape recovery. Basically, they recover ``static'' object for each time using multiple cameras.

In this paper, we propose simultaneous shape and motion recovery method without using explicit corresponding points.

The proposed method carves a high dimensional space which represents motion as well as shape. As a result, we can accomplish recovery of 3D scene which includes moving objects without using explicit corresponding points.

II.   EXTENDED MULTIPLE VIEW GEOMETRY

2.1 Projection from high dimensional space to image plane

At first, we summarize extended multiple view geometry[3]. By using the geometry, object shapes and motions can be recovered simultaneously as in the case of the existing 3D shape recovery method. In this section, we describe a camera projection from 4 or 5-dimensional space to 2-dimensional space by extended projective camera matrices.

At first, we describe a point projection from 4-dimensional space to image plane. A point $\mathbf{W} = [X \;\; Y \;\; Z \;\; \lambda]^T$ in 4-dimensional space is projected to 2-dimensional point $\mathbf{m} = [x \;\; y]^T$ as follows:

$$k\widetilde{\mathbf{m}} = \mathbf{Q}\widetilde{\mathbf{W}} \qquad (1)$$

where k is non-zero scalar, and $\widetilde{\mathbf{W}}$ and $\widetilde{\mathbf{m}}$ are homogeneous representations of $\mathbf{W}$ and $\mathbf{m}$. The matrix $\mathbf{Q}$ has 3×5 components and it indicates projection from 4D to 2D.

Next, we describe a projection from 5D to 2D. The projection from a point $\mathbf{V} = [X \;\; Y \;\; Z \;\; \lambda \;\; \mu]^T$ in 5D space to an image point $\mathbf{m}$ is described as follows:

$$l\widetilde{\mathbf{m}} = \mathbf{R}\widetilde{\mathbf{V}} \qquad (2)$$

where l is non-zero scalar and $\widetilde{\mathbf{V}}$ is homogeneous representation of $\mathbf{V}$. The matrix $\mathbf{R}$ has 3×6 components and it indicates projection from 5D to 2D.

Projections from high (*N*) dimensional space to image plane are accomplished by $3 \times (N+1)$ matrix. We call the matrix extended projection matrix, and we call multiple view geometries for high dimensional space extended multiple view geometry. The extended multiple view geometry can describe various relationships which cannot be described by existing multiple view geometry for 3-dimensional space. The extended multiple view geometry can describe dynamic configurations as shown in the following sections.

2.2 Projection of 1D linear moving points

Let us consider a projection of 3D point $\mathbf{X}$ moving in the

direction of $\mathbf{U}$ with speed λ. The position of $\mathbf{X}$ at time t can be represented by $\mathbf{X} + \lambda t\mathbf{U}$. Now, let us consider the case where the point $\mathbf{X}$ is projected to image plane by camera $\mathbf{P}$. The projection of $\mathbf{X}$ is described by:

$$k\tilde{\mathbf{m}} = \mathbf{P}(\mathbf{X} + t\lambda\mathbf{U}) \tag{3}$$

where k is non-zero scalar. Now, let us define a 3×5 extended projection matrix $\mathbf{Q}$ which represents projection from 4D point to image plane, and a 4-dimensional point $\mathbf{W}$ as follows:

$$\mathbf{Q} = \begin{bmatrix} \mathbf{P} & t\mathbf{PU} \end{bmatrix}, \tag{4}$$

$$\mathbf{W} = \begin{bmatrix} \mathbf{X}^T & \lambda \end{bmatrix}^T, \tag{5}$$

By using $\mathbf{Q}$ and $\mathbf{W}$, Eq.(3) can be rewritten as follows:

$$k\tilde{\mathbf{m}} = \mathbf{Q}\tilde{\mathbf{W}} \tag{6}$$

where $\tilde{\mathbf{W}}$ is a homogeneous representation of $\mathbf{W}$. This equation indicates that the projection of a moving point in 3D space can be described as a projection of a static point in 4-dimensional space. Thus, we can recover shape $\mathbf{X}$ and motion λ from input images.

2.3 Projection of planar moving points

Next, we discuss a projection of planar moving point in 3D space. Let us consider a 3D point $\mathbf{X}$ which moves on a plane in 3D space. If we consider two independent vectors $\mathbf{U}_1$ and $\mathbf{U}_2$ on the plane, the moving point at time t can be described as $X + t\lambda\mathbf{U}_1 + t\mu\mathbf{U}_2$, where λ and μ indicate speed of point $\mathbf{X}$ in each direction.

Now, let us consider a projection of the moving point by camera $\mathbf{P}$. The projection of the moving point is described by:

$$l\tilde{\mathbf{m}} = \mathbf{P}(\mathbf{X} + t\lambda\mathbf{U}_1 + t\mu\mathbf{U}_2) \tag{7}$$

We can also consider this projection in high dimensional space. Let us define a 3×6 extended projection matrix $\mathbf{R}$ and 5-dimensional vector $\mathbf{V}$ as follows:

$$\mathbf{R} = \begin{bmatrix} \mathbf{P} & t\mathbf{PU}_1 & t\mathbf{PU}_2 \end{bmatrix}, \tag{8}$$

$$\mathbf{V} = \begin{bmatrix} \mathbf{X}^T & \lambda & \mu \end{bmatrix}^T, \tag{9}$$

By using $\mathbf{R}$ and $\mathbf{V}$, Eq.(7) can be rewritten as follows:

$$l\tilde{\mathbf{m}} = \mathbf{R}\tilde{\mathbf{V}} \tag{10}$$

where $\tilde{\mathbf{V}}$ denotes homogeneous representation of $\mathbf{V}$. The Eq.(7) indicates that the projection of a planar moving point in 3D space can be considered as a projection of a static point in 5D space.

The Eqs.(3) and (7) indicate that projections of moving points in 3D space can be considered as projections of static points in high dimensional space, which include shape and planar motion information. Furthermore, the equations also indicate that we can recover not only object shape, but also motion from image points.

Note that, the theory can also represent arbitrary 3D motion using a projection from 6-dimensional space[3].

2.4 Multilinear constraint on high dimensional spaces

In the section, we describe multiple view constraints from high dimensional spaces to image space. Now, let us consider a case where an N-dimensional point $\mathbf{V}$ is projected onto image points $\mathbf{m}_1, \mathbf{m}_2, \cdots, \mathbf{m}_M$ by M extended cameras represented by $3 \times (N+1)$ matrices $\mathbf{R}_1, \mathbf{R}_2, \cdots, \mathbf{R}_M$. The projections can be described as follows:

$$l_1\tilde{\mathbf{m}}_1 = \mathbf{R}_1\tilde{\mathbf{V}}$$
$$l_2\tilde{\mathbf{m}}_2 = \mathbf{R}_2\tilde{\mathbf{V}}$$
$$\vdots \tag{11}$$
$$l_M\tilde{\mathbf{m}}_M = \mathbf{R}_M\tilde{\mathbf{V}}$$

where l_i is non-zero scalar. The equations can be rewritten as follows:

$$\begin{bmatrix} \mathbf{R}_1 & \tilde{\mathbf{m}}_1 & \mathbf{0} & & \mathbf{0} \\ \mathbf{R}_2 & \mathbf{0} & \tilde{\mathbf{m}}_2 & \cdots & \mathbf{0} \\ & & \vdots & & \\ \mathbf{R}_M & \mathbf{0} & \mathbf{0} & & \tilde{\mathbf{m}}_M \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{W}} \\ l_1 \\ l_2 \\ \vdots \\ l_M \end{bmatrix} = \mathbf{0} \tag{12}$$

The equation indicates that the matrix in left term has a null space. Thus, the following equation holds.

$$\det \begin{bmatrix} \mathbf{R}_1 & \tilde{\mathbf{m}}_1 & \mathbf{0} & & \mathbf{0} \\ \mathbf{R}_2 & \mathbf{0} & \tilde{\mathbf{m}}_2 & \cdots & \mathbf{0} \\ & & \vdots & & \\ \mathbf{R}_M & \mathbf{0} & \mathbf{0} & & \tilde{\mathbf{m}}_M \end{bmatrix} = \det \mathbf{M} = 0 \tag{13}$$

We call the constraint multilinear constraint. For example, the multilinear constraint for three cameras in 5D space is described as follows:

$$\det \begin{bmatrix} \mathbf{R}_1 & \tilde{\mathbf{m}}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{R}_2 & \mathbf{0} & \tilde{\mathbf{m}}_2 & \mathbf{0} \\ \mathbf{R}_3 & \mathbf{0} & \mathbf{0} & \tilde{\mathbf{m}}_3 \end{bmatrix} = \det \mathbf{M}_5 = 0 \tag{14}$$

where $\mathbf{R}_i$ indicates an extended projection matrix for i-th camera. The cofactor expansion of Eq.(14) derives the following equation.

$$m^i m'^j m''^k T_{ijk} = 0 \tag{15}$$

where m, m' and m'' indicate tensor notations of $\tilde{\mathbf{m}}_1, \tilde{\mathbf{m}}_2$ and $\tilde{\mathbf{m}}_3$. The T denotes trifocal tensor which represents relative relationship among three cameras. The equation indicates that we can calibrate trifocal tensor T among three cameras from only image corresponding points. Furthermore, we can estimate an image point $\mathbf{m}_3$ from only $\mathbf{m}_1, \mathbf{m}_2$ and T without recovering 5-dimensional point explicitly.

III. SPACE-TIME CARVING FOR RECOVERY OF SHAPE AND MOTION

3.1 Space carving for 3D shape recovery

In Section II, we introduced extended multiple view geometries

in order to represent object shapes and motions simultaneously. Of course, we need to find corresponding points from input image set in order to recover shape and motion by using the geometries. In particular, if we would like to recover dense 3D shape, a lot of corresponding points are required. On the other hand, Kutulakos et al.[6] proposed space carving method which needs no explicit corresponding points for 3D shape recovery. By extending of the space carving using extended multiple view geometry, we propose space-time carving which can recover not only shape but also motion. In order to derive space-time carving, we summarize the existing space carving method in this section.

Let us consider 2 cameras taking the same scene from different points at the same time. Now, we assume that a surface of the object is Lambertian surface. In this case, when the two cameras observe the same 3D point, observed colors in each camera is the same. On the other hand, the colors are not same when two cameras observe different 3D points. We call the property ``photo consistency'' in this paper.

Let us consider the point $\mathbf{m}_1$ in one of the camera image and point $\mathbf{m}_2$ in the other camera image. The photo consistency $p(\mathbf{m}_1,\mathbf{m}_2)$ is defined as follows:

$$p(\mathbf{m}_1,\mathbf{m}_2) = \begin{cases} 1 & \text{if } \| \mathbf{c}(\mathbf{m}_1) - \mathbf{c}(\mathbf{m}_2) \| < \theta_p \\ 0 & \text{otherwise} \end{cases} \quad (16)$$

where $\mathbf{c}(\mathbf{m})$ indicates color vector of point $\mathbf{m}$, and θ_p is a threshold of color difference.

Then, we define a space carving rule S_3 for a 3D point $\mathbf{X}$ by using photo consistency p as follows:

$$S_3(\mathbf{X}) = p(\mathbf{P}_1\tilde{\mathbf{X}}, \mathbf{P}_2\tilde{\mathbf{X}}) \quad (17)$$

where $\mathbf{P}_1$ and $\mathbf{P}_2$ are projection matrix for each camera. If $S_3(\mathbf{X})=0$, the point $\mathbf{X}$ is carved, and if $S_3(\mathbf{X})=1$, the point $\mathbf{X}$ is left in the space.

The recovery of 3D shape is realized by the following steps using the carving rule.
1) Given 3D space is separated into set of voxels.
2) Each voxel $\mathbf{X}_i$ is carved or left according to the by evaluation of $S_3(\mathbf{X}_i)$
3) After the all voxel evaluations, the remaining voxels represent the 3D surface of the scene.

Let us consider a scene indicated in Fig.1 for detailed explanation of space carving. We consider about three points $\mathbf{X}_a$ onto an object surface, $\mathbf{X}_b$ far from an object and $\mathbf{X}_c$ nearby object. For a point $\mathbf{X}_a$, a photo consistency $p(\mathbf{P}_1\tilde{\mathbf{X}}_a, \mathbf{P}_2\tilde{\mathbf{X}}_a)$ is valid and the point is left in the space. Of course a photo consistency of $\mathbf{X}_b$ is not valid and the point is carved. Unfortunately, a photo consistency of $\mathbf{X}_c$ is valid in spite of the point is not located onto the object surface because observed colors of $\mathbf{X}_c$ by camera 1 and camera 2 is similar. And then, a recovery shape by space carving is larger than truth object. The recovered shape by the method is called as ``photo-hull''.
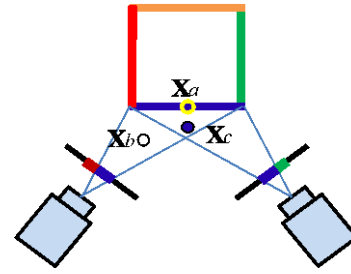

Fig. 1. Space carving by using photo consistency.

The space carving method can recover 3D scene without explicit corresponding points. We need only images taken by cameras and relationship such as fundamental matrices for the scene recovery. In the following sections, we extend this method for recovery not only shapes but also motions by using extended multiple view geometry.

Note that, although we assume that a surface of an object is Lambertian surface in this paper, we may use another reflectance model for photo consistency such as Phong model[7], Torrance-Sparrow model[8] and so on.

3.2 Shape and motion recovery by carving of high dimensional spaces

Now, we propose method which can recover shape and motion simultaneously without explicit corresponding points in input images. The method is based on the space carving method described in previous section. The space carving method carves 3-dimensional space which represents 3D shapes for shape reconstruction. On the other hand, our proposed method carves high dimensional space which describes not only shape, bus also object motions such as described in *2.2* and *2.3*. The space is constructed by components of space and time, and then, the space can be regarded as space-time [9]. Thus, we call the carving method space-time carving.

At first, we described carving of 4-dimensional space which is described in *2.2*. Now, let us consider 2 extended projection matrix $\mathbf{Q}_1$ and $\mathbf{Q}_2$ observing target scene. A linear moving point $\mathbf{X}$ with speed λ is denoted by static point $\mathbf{W}$ in 4-dimensional space which represents shape and motions. A point $\mathbf{W}$ is carved or left in the space by a following rule:

$$S_4(\mathbf{X}) = p(\mathbf{Q}_1\tilde{\mathbf{W}}, \mathbf{Q}_2\tilde{\mathbf{W}}) \quad (18)$$

Since the point $\mathbf{W}$ includes shape and motion information, we recover shape and motion simultaneously without explicit corresponding point in the input images.

A carving of 5-dimensional space which includes 3D shape and planar motion can also be accomplished. When a point $\mathbf{V}$ in 5D space is projected by extended projection matrix $\mathbf{R}_1$ and $\mathbf{R}_2$, 5D space recovery is realized by the following rule:

$$S_5(\mathbf{X}) = p(\mathbf{R}_1\tilde{\mathbf{V}}, \mathbf{R}_2\tilde{\mathbf{V}}) \quad (19)$$

Since the point $\mathbf{V}$ includes shape and 2-dimensional motion information, we can recover shape and planar motion information simultaneously without using explicit corresponding points.

3.3 High dimensional space recovery under weak calibrations

In Section *3.2*, we proposed simultaneous shape and motions recovery by carving of high dimensional spaces. The method, however, requires full calibration of extended projection matrices. In this section, we propose a method which can be used even when we know only relative relationship between cameras such as trifocal tensors.

Now, let us consider 4-dimensional space recovery, where we calibrate only trifocal tensor which represents relative relationship among 3 extended projection matrices. The space time carving based on the trifocal tensor is realized by the following steps. Let $\mathbf{m}_1, \mathbf{m}_2$ and $\mathbf{m}_3$ denote image point of first, second and third image.

1) We select arbitrary points $\mathbf{m}_1$ and $\mathbf{m}_2$ from first and second images.

2) If $p(\mathbf{m}_1, \mathbf{m}_2) = 1$, an image point $\mathbf{m}_3$ which corresponds to $\mathbf{m}_1$ and $\mathbf{m}_2$ is estimated by Eq.(15).

3) If $p(\mathbf{m}_1, \mathbf{m}_3) = 1$, a 4-dimensional point $\mathbf{W}'$ is recovered projectively from points $\mathbf{m}_1$, $\mathbf{m}_2$ and $\mathbf{m}_3$.
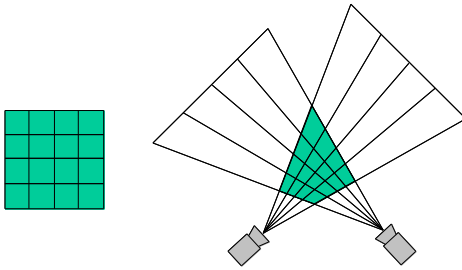

Fig. 2. Orthogonal grid voxels (left) and projective grid voxels (right).

Fig. 2 (left) indicates the voxel space used by the method described in section *3.2*, and Fig.2 (right) indicates the voxel space used in the method described in this section. The right voxel space has projective ambiguity and the space is given by projective transformation of left set. Thus, the recovered point $\mathbf{W}'$ includes projective ambiguity. However, we do not need to eliminate the ambiguity because there are various applications which can be realized from projective reconstruction [9].

A recovery of 5-dimensional space can be realized by the same way. For the recovery of 5-dimensional space, 4D point $\mathbf{W}'$ is replaced to 5D point $\mathbf{V}'$ and it includes object shape and planar motion information of the 3D point. In this case, the recovered point $\mathbf{V}'$ includes 5-dimensional projective ambiguity.

Note that, we can eliminate the *N*-dimensional projective ambiguity by a projective transformation. The projective transformation is accomplished by the following computation.

$$k\hat{\mathbf{W}} = \mathbf{H}\widetilde{\mathbf{W}} \qquad (20)$$

where $\mathbf{H}$ is $(N+1) \times (N+1)$ projective transform matrix and $\hat{\mathbf{W}}$ and $\mathbf{W}$ are real vector and recovered vector. Since the matrix has $(N+1) \times (N+1) - 1$ DoF and Eq.(20) gives N linear independent equations, the matrix can be estimated from more than $((N+1) \times (N+1) - 1)/N$ which has known shape and motion.

By transforming using the estimated matrix, projective ambiguity of recovered vectors are eliminated.

Furthermore, we can eliminate ambiguity of arbitrary components in recovered vector. For example, we consider elimination of motion ambiguity λ from 4-dimensional vector $[X \quad Y \quad Z \quad \lambda]^T$. When 5×5 matrix $\mathbf{H}_5 = [\mathbf{h}_1 \quad \mathbf{h}_2 \quad \mathbf{h}_3 \quad \mathbf{h}_4 \quad \mathbf{h}_5]$ denote a 4-dimensional projective transform matrix to real vectors, the elimination ambiguity is accomplished by

$$k\begin{bmatrix} \hat{\lambda} \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{h}_4^T \\ \mathbf{h}_5^T \end{bmatrix} \widetilde{\mathbf{W}} \qquad (21)$$

where $\hat{\lambda}$ is a real motion scalar. Since the matrix $[\mathbf{h}_4 \quad \mathbf{h}_5]^T$ has 9 DoF and the equation gives an independent equation, we can estimate the transform matrix from 9 corresponding points. Note that, we do not need shape information for the estimation; we need only motion information of the object.

IV.   EXPERIMENTAL RESULTS

4.1 Recovery of planar moving objects using real images

In this section, we show some experimental results by our proposed method. At first, we show object motion recovery result from real images. In this experiment planar motion objects are taken by a camera. Note that, we does not need multiple cameras for shape and motion recovery since Eq.(9) indicates that static camera in multiple instants can be regarded as multiple cameras in 5-dimensional space. The cameras are completely calibrated and the space-time carving method described in *3.2* was used.

The size of target 3D space is $40cm \times 30cm \times 20cm$ and basis vectors $\mathbf{U}_1 = [10cm \quad 0cm \quad 0cm]$ and $\mathbf{U}_1 = [0cm \quad 0cm \quad 10cm]$ are used for motions. The size of motions for $\mathbf{U}_1$ and $\mathbf{U}_2$ are $-20cm/$sec and $20cm/$sec, respectively. The size of each voxel I $0.1cm \times 0.1cm \times 0.1cm \times 10cm/\sec \times 10cm/\sec$. An experimental environment and target objects are shown in Fig. 3.
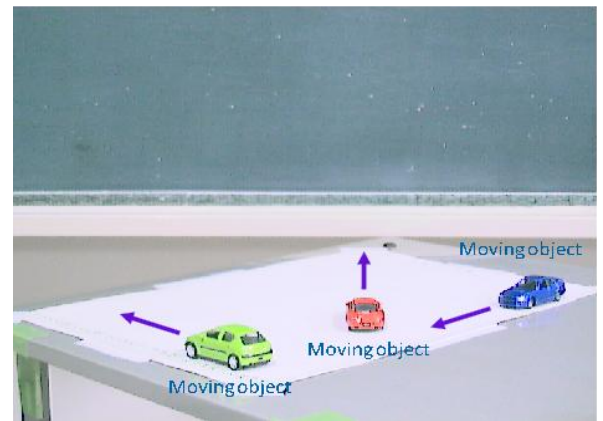

Fig. 3. Experimental environments and moving direction of objects.

The target objects were moved to a direction indicated in the image with constant velocities. The target objects are taken three times by a static camera and the shape and motions are

recovered from the three images. The taken images are shown in Fig. 4 which size is 640×480. In the images, target objects are extracted beforehand by a simple background subtraction.

The object motion recovery result is shown in Fig.5. In the image, recovered voxels are the color of the voxels indicates direction magnitude of motions. A color circle in left bottom shows direction and speed for each color. A hue of color indicates direction and saturation indicates speed of the motions. For example, if an object moves to right, the object is colored by yellow. The direction and speed estimated by the proposed method is correct in most of the case, and then, the results indicate that our proposed method works well.

Fig.4. Examples of input images.
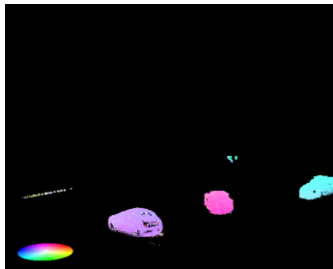
Fig. 5. Object motion recovery result by our proposed method.

4.2 Motion recovery under weak calibration

In this experiment, we recovered shape and motion under weak calibration. Objects in the experiment moves to the same direction and the shape and motion can be represented in 4-dimensional space. The motion of the objects is indicated in Fig. 6 and input images are indicated in Fig. 7.
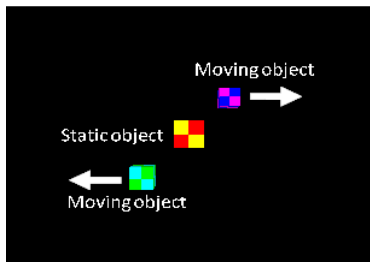

Fig. 6. Motions of target objects.

In this experiment, only a trifocal tensor which represents relative relationship among camera is calibrated. By using the shape and motion recovery method described in *3.3*, the information and recovered, and then each pixel which are reprojections of recovered information are colored by motion information.

An upper image in Fig. 8 indicates recovered motion information. Color bars in the images indicate direction and speed for each pixel. In the figure, objects which move to

right/left become green/purple, and static objects become grey. A lower image in the figure indicates estimation errors. In this figure, each pixel is colored by information of estimation error, and then, pixels become grey (which is center of color bar) when estimation error is small. All objects in the result image is colored by grey image, and then, the results indicate that our proposed method can recover motion information under weak calibration.
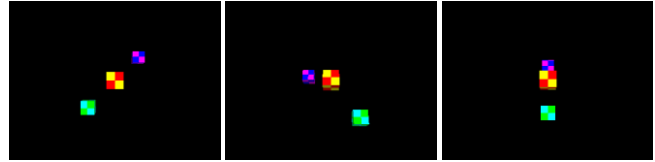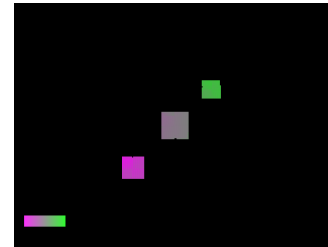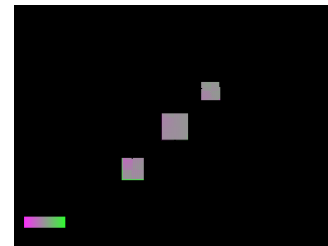

Fig. 7. Examples of input images.

Recovered motion

Motion estimation error

Fig. 8. Motion recovery result (upper row) and estimation error (lower row).

4.3 Recovery of non-rigid motion

In this experiment we recover motion of non-rigid object. In this scene, the shape of a 3D ball was transformed as shown in Fig.9. The motion of each point is recovered by the method described in *3.2* and each point is colored according to the recovered motion information. The size of 3D scene is $5cm \times 5cm \times 5cm$ and basis vectors $\mathbf{U}_1 = \begin{bmatrix} 0.5cm & 0cm & 0cm \end{bmatrix}$ and $\mathbf{U}_2 = \begin{bmatrix} 0cm & 0.5cm & 0cm \end{bmatrix}$ are used for motions representation. The size of each voxel is
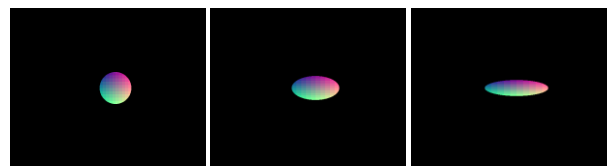$0.1cm \times 0.1cm \times 0.1cm \times 0.1cm / \sec \times 0.1cm / \sec$.

Fig. 9. Input images for non-rigid motion recovery.

Recovered motion

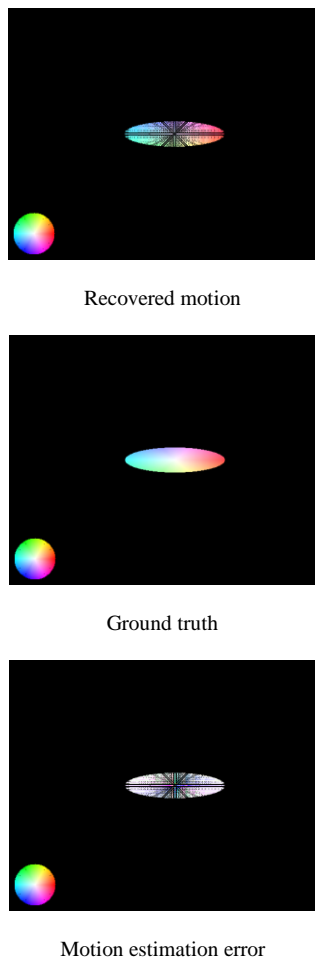Ground truth

Motion estimation error

Fig. 10. Non-rigid motion recovery results.

Fig. 10 shows recovered motion information, ground truth of motions and difference from ground truth. The recovered result is mostly identical to the ground truth of the motions. The results indicate that our proposed method recovers motions even if the object motion is not rigid motion.

V.   CONCLUSIONS

In this paper, we proposed a novel shape and motion recovery method which requires no corresponding points. The proposed method carves high dimensional spaces which represents not only 3D shape but also 3D motions, and thus, the method can recover 3D scene which includes moving objects without explicit corresponding points. Furthermore, we showed that the proposed method can recover the scenes in the case where only weak calibration is available. Finally, some experimental results indicate that the proposed method can be applied to rigid and non-rigid object motions recovery.

REFERENCES

[1]   R. Hartley, A. Zisserman. *Multiple View Geometry in Computer Vision. Cambridge University Press*，2000.
[2]   O. Faugeras, R. Keriven. Scale-space and affine curvature. In *Proc. Europe-China Workshop on Geometrical Modeling and Invariants for Computer Vision*, pp. 17—24, 2001.

[3]   L. Wolf, A. Shashua.: On projection matrices $p^k \rightarrow p^2, k = 3, \cdots, 6$, and their applications in computer vision. *International Journal of Computer Vision* vol.48, no**.**3, pp. 53-66, 2002.
[4]   L. Carceroni, K. Kutulakos. Multi-view 3d shape and motion recovery on the spatio-temporal curve manifold. In Proc. *the Seventh IEEE International Conference on Computer Vision* (ICCV1999), pp. 520-527, 1999.
[5]   S. Vedula, S. Baker, S. Seitz, T. Kanade. Shape and motion carving in 6d. In Proc. *IEEE Conference on Computer Vision and Pattern Recognition* (CVPR2000), pp. 592-598, 2000.
[6]   K. Kutulakos, S. Seitz. A theory of shape by space carving. *International Journal of Computer Vision* vol.38, no.3, pp. 199-218, 2000.
[7]   B. Phong. *Illumination for computer generated pictures. Communications of the ACM vol.*18, pp. 311-317 , 1975.
[8]   K. Torrance, E. Sparrow. *Theory for off-specular reflection from roughened surface. J. Opt. Soc. Am vol.*57 , pp.1105-1114, 1967.
[9]   K. Hayakawa, J. Sato. Multiple view geometry in the space-time. In Proc. *Asian Conference on Computer Vision.* Vol.~2, pp. 437—446**,** 2006.

Fumihiko Sakaue received the B.E., M.E. and Ph D. degrees in Information technology from Okayama University, Japan in 2001, 2003 and 2006, respectively. He is currently a   research assistant professor at the Department of Computer Science and Engineering, Nagoya Institute of Technology.

   His research interests include computer vision, photometric image analysis, pattern recognition and mixed reality.
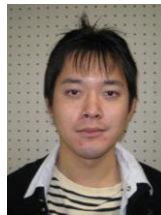
Masaaki Takami received B.E. and M.E. degree from the Department of Computer Science, Nagoya Institute of Technology, Japan in 2007 and 2009, respectively.

   His research interests include computer vision, 3D reconstruction and Photometry.

Jun Sato received B.E. degree from Nagoya Institute of Technology, Japan in 1984, and Ph.D. degree in information engineering from the University of Cambridge, England in 1997. From 1996 to 1998 he was a research associate at the Department of Engineering, University of Cambridge. During 1997--1998, he was an invited researcher at ATR Human Information Processing Research Laboratories, Kyoto. He joined the Department of Electrical and Computer Engineering, Nagoya Institute of Technology in 1998 as an associate professor. He is currently a professor at the Department of Computer Science and Engineering, Nagoya Institute of Technology.

   His research interests include computer vision, geometric invariants, visual navigation, visual interface and mixed reality. He was awarded the Best Science Paper Prize of BMVC twice in 1994 and 1997, the Best Paper Prize of SSII'99 and the Special Session Prize of MIRU'07, etc. He is a member of the IEEE, IEICE and IPSJ.