

# Entire Shape Acquisition Technique Using Multiple Projectors and Cameras with Parallel Pattern Projection

RYO FURUKAWA<sup>1,a)</sup> RYUSUKE SAGAWA<sup>2</sup> HIROSHI KAWASAKI<sup>3</sup>  
KAZUHIRO SAKASHITA<sup>4</sup> YASUSHI YAGI<sup>4</sup> NAOKI ASADA<sup>1</sup>

Received: April 15, 2011, Accepted: October 12, 2011, Released: March 28, 2012

**Abstract:** In the present paper, we propose a one-shot scanning system consisting of multiple projectors and cameras for dense entire shape acquisition of a moving object. One potential application of the proposed system is to capture a moving object at a high frame rate. Since the patterns used for one-shot scanning are usually complicated, and the patterns interfere with each other if they are projected onto the same object, it is difficult to use multiple sets of patterns for entire shape acquisition. In addition, the overlapped areas of each object have gaps and errors are accumulated. As such, merged shapes are usually noisy and inconsistent. In order to address this problem, we propose a one-shot shape reconstruction method using a projector to project a static pattern of parallel lines of one or two colors. Since each projector projects only parallel lines with a small number of colors, these patterns are easily decomposed and detected even if the patterns are projected multiple times onto the same object. We also propose a multi-view reconstruction algorithm for the projector-camera system. In the experiment, we built a system consisting of six projectors and six cameras, and dense shapes of entire objects were successfully reconstructed.

**Keywords:** shape reconstruction, entire shape acquisition, line pattern, de Bruijn sequence, projector-camera system

## 1. Introduction

Dense entire shape acquisition of moving objects is required in a number of fields. For example, in order to visually capture a dancing person wearing soft clothing, an entire shape acquisition method with high resolution and high speed is preferable. Thus far, for the purpose of capturing the entire shape of such moving objects, the shape from silhouette [1] technique has been widely used. Although this technique is suitable for capturing moving objects, concavity and small bumps are difficult to retrieve using this technique. In terms of passive systems such as shape from silhouette (i.e., a system using only cameras), the multi-view stereo (MVS) technique has been researched extensively with the goal of recovering the 3D shape of static objects, and the results have improved greatly in recent years. Although several attempts have been made to extend the MVS technique to capturing moving objects, recovering dense and precise shapes of dynamic scenes using the MVS technique remains a difficult task, especially, if the texture is uniform and the number of images is small.

On the other hand, active 3D scanning methods are used for practical purposes, because such methods can capture the object with high accuracy and robustness. In particular, a great deal of research and development has been conducted on structured light systems that consist of a projector and a camera because such

systems are simple and yet can capture textureless objects.

In terms of the projector-camera-based structured light systems, there are two types of systems, temporal-encoding-based systems and spatial-encoding-based systems. Temporal-encoding-based systems use “time-multiplexing strategy” [2]. In this strategy, a set of patterns are projected onto the target, the images are captured for each of the patterns, and mapping of correspondences from the camera image to the projector image is reconstructed from the sequence of illuminations for each camera pixel. On the other hand, spatial-encoding-based systems use information of “spatial neighborhood” [2] of each point of the pattern to identify the point observed from the camera image. Since only a single image of pattern is needed to reconstruct correspondence mapping in this strategy, it is suitable to capture dynamic scenes or objects.

Although these techniques are well suited to a single system used for capturing one side of the targets, several fundamental problems remain when applying these techniques to multiple sets of systems surrounding a target object in order to capture the entire shape of the object. For a temporal-encoding-based system, since a set of images is required for reconstruction, it is difficult to achieve entire shape acquisition in a short time. On the other hand, for a system based on a spatial-encoding technique, since the pattern used for the system is usually complicated and the patterns interfere with each other if they are projected onto the same object, it is difficult to use multiple sets of systems to capture the entire shape of an object. Furthermore, since the errors for each scan are accumulated, large gaps occur when the obtained results are integrated.

<sup>1</sup> Hiroshima City University, Hiroshima 731–3166, Japan

<sup>2</sup> National Institute of Advanced Industrial Science and Technology, Tsukuba, Ibaraki 305–8568, Japan

<sup>3</sup> Kagoshima University, Kagoshima 890–8580, Japan

<sup>4</sup> Osaka University, Ibaraki, Osaka 567–0047, Japan

<sup>a)</sup> ryo-f@hiroshima-cu.ac.jp

If a spatial-encoding-based projector-camera system using a small number of colors with a parallel line pattern is realized, then the aforementioned line detection problem can be resolved. Recently, methods that can reconstruct the shape from a single image using a static grid pattern with only a single color or two colors have been proposed [3], [4], [5]. However, since the projected pattern consists of vertical and horizontal dense parallel lines, it is still difficult to decompose these patterns when they are projected from multiple projectors onto the same object. Furthermore, since these methods use the density of line intervals to retrieve a unique solution, it is impossible to make both vertical and horizontal lines dense, which results in not only sparse reconstruction, but also in frequent failure in reconstruction of small objects.

In the present paper, we propose a shape reconstruction method using intersections of two sets of parallel lines from two projectors. Compared to the patterns of previous spatial-encoding-based projector-camera systems where the patterns are usually complicated with a number of colors, the patterns used in the proposed method is much simpler. Thus, interference of patterns is greatly reduced. A similar technique using a grid pattern for a single projector is proposed [4]. However, simple increment of the number of the projector drastically increase the complexity of the pattern by overlapping grid patterns. Because of the advantage of the proposed method that detection of the patterns is not severely affected by overlapping them, freedom on installation of multiple projectors is significantly improved and entire shape acquisition becomes possible.

Another advantage of the proposed method is that a unique solution can be retrieved using only the intersection points. Thus, both the vertical and horizontal lines can be as dense as possible, allowing the reconstruction of small objects. In addition, we propose a multi-view reconstruction algorithm that uses shape reconstruction based on grid patterns [4]. In the experiments described later, the effectiveness and accuracy of the proposed technique are evaluated and confirmed using synthetic data. In addition, we constructed the actual system that consists of multiple projectors and cameras in order to compare the proposed method to other methods and demonstrate its advantages.

The main contributions of the present paper are as follows: (1) one-shot dense entire shape reconstruction using multiple projectors and cameras is proposed, (2) a linear solution for shape reconstruction from a grid pattern is presented, (3) a reconstruction algorithm for a multi-view projector-camera system is presented, and (4) an actual system that can capture the entire shape of an object is constructed.

## 2. Related Research

Entire shape acquisition systems have been researched extensively for various applications, e.g., CAD and motion capture, and a large number of techniques based on multi-view stereo (MVS) have been proposed [6], [7], [8]. However, if the texture information is insufficient and the number of input images is small, then the density of the reconstruction will become sparse and its accuracy become low. In addition, the number of parameters to be optimized and the high computational cost are prob-

lematic in actual systems. Recently, several research groups have applied MVS to dynamic scenes [9], [10], [11]. However, these methods are still under development and require further research before practical application. In terms of multiple-camera systems, owing to their calculation efficiency and their robustness, silhouette-based techniques are commonly used to capture dynamic objects [12], [13]. However, these techniques still have several critical problems, such as sparse and inaccurate reconstruction, especially for textureless and concave areas of objects.

Thus far, for practical 3D shape acquisition purposes, active measurement systems have been widely used because they are stable even if there are no textures or features on the target scenes. Some of them are based on illumination by laser light sources. For instance, laser triangulation scanners [14] uses line-laser light source and measure the scenes by light sectioning method (triangulation between the laser-swept plane and the line of sight). Time-of-flight laser scanners [15] measures the distance to each of the laser-spotted points by measuring the time that a light goes from the light source to the point and returns back. Currently, time-of-flight one-shot cameras are also available [16], [17] based on observing modulated illumination by infra-red LEDs. The above systems needs specialized hardware devices, thus it is not easy to use multiple set of them to capture entire shapes because of costs and technological problems such as synchronization, and there are limits of the speeds of the target object because they need time for scanning the light sources or observing modulation of illuminations.

A large category of active measurement systems is projector-camera-based 3D measurement systems [18], [19]. In terms of these systems, techniques for reconstructing 3D information can be categorized into two types, namely, temporal-encoding techniques and spatial-encoding techniques. Although temporal-encoding techniques can achieve dense and accurate reconstructions, these techniques require a set of images in which the object is captured using different projected patterns. Thus, it is difficult to use such techniques to reconstruct moving objects. In addition, it is impossible to use multiple sets of them at the same time. Douglas et al. proposed a temporal-encoding-based entire shape acquisition system by using a special set-up of multiple mirrors [20]. However, it is also difficult to extend this technique to general applications.

On the other hand, techniques that use only spatial encoding of a pattern, allowing scanning using only a single-frame image, have been proposed [21], [22]. Since such techniques require only a single image for reconstruction, there is a good possibility to scan moving objects. However, these techniques typically use complex patterns with several colors for decoding and may be easily affected and leads to ambiguities on textured object or near depth boundaries. In addition, if these patterns are projected onto the same object for entire shape acquisition, then they interfere with each other, making it difficult to decompose them, even if an efficient algorithm is applied [4].

Recently, solutions for the complex pattern of the spatial-encoding method using a simple grid pattern that embeds information in relation of connection of parallel lines have been published [3], [4], [5]. However, since the system still projects dense

vertical and horizontal grid patterns, it remains difficult to decompose these patterns after they have been projected onto the same object. If the pattern consists strictly of one-directional parallel lines with one or two colors for each projector, then the problems can be greatly simplified. We propose such a technique in the present paper.

In space-time stereo [18], [23], [24], the correspondence problem of stereo method using two cameras (i.e., passive stereo) is solved using active illumination by a projector. In order to capture highly dynamic scenes using this method, either the pattern should be static, or high-speed synchronization is needed. High-speed synchronization with a large number of camera is difficult. About using static patterns with MVS, a large complexity of such system is a problem, because a large number of frames should be reconstructed to capture a dynamic scene. As far as our knowledge, there have been no works that entire shape acquisition is realized using space-time stereo.

Several studies have focused on MVS for projector-and-camera systems, especially for temporal-coding-based systems [25]. However, applying conventional MVS to grid-pattern-based systems is not easy, because point correspondences are not explicitly provided. Therefore, a solution to this problem is required.

### 3. Dense Entire Shape Acquisition System

#### 3.1 System Configuration

In the proposed 3D measurement system, a camera observes line patterns projected from multiple projectors. The projected pattern is fixed and does not change, so synchronization is not required. The minimum setup consists of two projectors and a camera, as shown in Fig. 1. The camera and the projector are assumed to be calibrated (i.e., the intrinsic parameters of the devices and their relative positions and orientations are known). The projectors project vertical and horizontal line patterns and the system reconstructs the shape of the target object by observing the intersection points of the lines. Although the proposed method can reconstruct shapes from monochrome line patterns, we use colored patterns to improve the accuracy and robustness. In the present paper, a periodic pattern of two colors based on the de Bruijn sequence is used.

Each line encoded by the de Bruijn sequence can be identified in the cycle of the pattern by observing the local subsequence. The detailed definition of pattern is described in Section 3.3.

When we reconstruct the entire shape of an object, multiple projectors and cameras are used, as shown in Fig. 2. The projectors and cameras are placed such that each camera can observe the intersection points of line patterns. In order to discriminate the patterns projected from different projectors, color information is used. In the present paper, we experimented with six projectors and six cameras.

#### 3.2 Algorithm Overview

An overview of the proposed algorithm is shown in Fig. 3. First, the vertical and horizontal projected patterns are detected separately, and the points of intersection of the lines on an object are found. Consequently, lines that form grid graphs in an image are obtained.

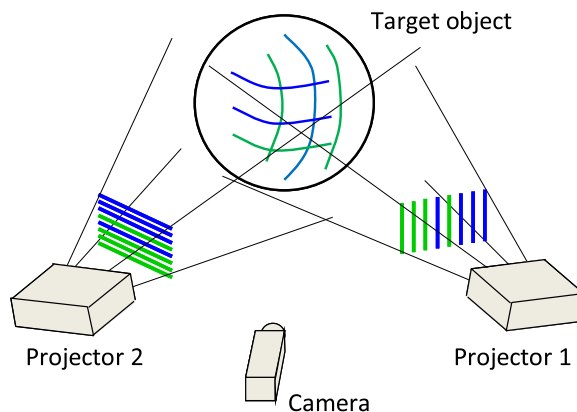


Fig. 1 The minimum setup of the proposed approach consists of two projectors and a camera. The projectors project vertical and horizontal line patterns and the camera observes the intersection points of the lines.

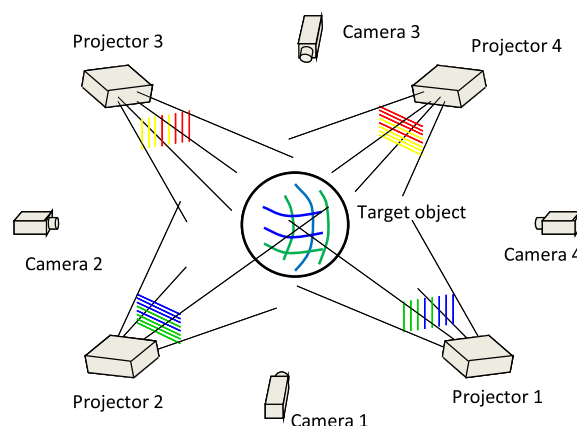


Fig. 2 An example setup to reconstruct the entire shape consisting of four projectors and four cameras. The patterns are constructed using different colors in order to discriminate the vertical and horizontal patterns.

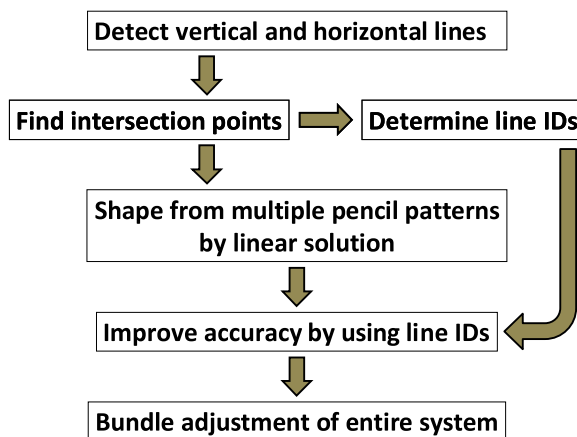


Fig. 3 Overview of the proposed algorithm.

Each of the lines projected from a projector sweeps a plane in the 3D space. A linear constraint of the plane parameters is obtained from the fact that an intersection point exists on two of the planes and the line of sight of the camera. By obtaining multiple constraints using the intersection points, a set of linear simultaneous equations of plane parameters can be constructed. By solving these simultaneous equations, the plane parameters are obtained.

In addition, in order to improve accuracy and computational efficiency, the local ID (i.e., an ID that is not globally unique) of

each line is obtained using color information (Section 3.3). The local IDs are used for the detection of incorrect connections between detected curves and for matching in subsequent steps.

Finally, 3D shapes are reconstructed by the light sectioning method using the estimated plane parameters. In reality, reconstruction sometimes becomes unstable because of the effect of noise or due to a limited number of intersection points, especially at a small area. Therefore, we propose a multi-view reconstruction technique in which reconstruction result is checked by observation of multiple cameras. Using the proposed method, the reconstruction errors of a unit system, which consists of two projectors and a camera, can be efficiently reduced.

### 3.3 Color Pattern and Line Detection

The proposed method uses a color code based on the de Bruijn sequence [21], [26], [27] to determine each line ID. A  $q$ -ary de Bruijn sequence of order  $n$  is a sequence of length  $q^n$  consisting of an alphabet of size  $q$ , in which every possible subsequence of length  $n$  is present exactly once. If a projected pattern is encoded by two or more symbols distinguished in a camera image, then the correspondence between an element in the projected pattern and the observed pattern is uniquely determined by matching subsequences of length  $n$  in a de Bruijn pattern.

Instead of using large  $q$  and  $n$  as some previous works [21], [27], we used periodic patterns generated with small  $q$  and  $n$ , as Sagawa et al. [4]. In the present paper, we used the number of colors  $q = 2$  and the length of codes  $n = 3$ . In other words, each cycle of the pattern consists of eight lines, and the line IDs are from 0 to 7.

In the case of  $q = 2$  and  $n = 3$ , a cycle of the de Bruijn sequence becomes AAABABBB, where A and B represent the two alphabets. If the pattern is encoded by using blue and green, all lines have blue components and the green components are off for A and on for B. Therefore, the blue component is used for line detection, and the alphabet of each line is determined by the green component.

The positions of the intersection points of the vertical and horizontal lines are computed with sub-pixel accuracy. In addition, for the case in which two intersection points that belong to different lines are incorrectly connected in the image, the incorrect connection can be cut by using the color pattern [4].

### 3.4 Calibration of a System with Multiple Projectors and Cameras

The proposed method determines shapes using multiple projectors and cameras. Therefore, to achieve metric reconstruction, both the cameras and the projectors must be intrinsically and extrinsically calibrated. In the present work, we assume that the cameras and projectors are intrinsically calibrated in advance, including lens distortions of the cameras. Many methods or codes can be found in public for this process [28], [29]. Then, in order to obtain extrinsic parameters and adjust focal lengths of the cameras and projectors, bundle adjustment technique is applied. By using bundle adjustment, all the information of correspondences are used simultaneously to avoid inconsistency between the devices. Since the correspondences of camera-camera, camera-

projector, and projector-projector are required for bundle adjustment, they are obtained as follows.

First, the corresponding points between a camera and a projector are obtained by projecting structured-light patterns onto the target object from a projector. We use time-multiplexed patterns based on Gray codes. In this step, the target object can be captured multiple times with different positions. In this case, the captured data can be treated as correspondence data of an union of the shapes of the target object with those multiple positions. This technique is useful if the target object can be captured only from specific sides (e.g., an one-sided plane).

Next, if the patterns of a projector are observed by two cameras, the correspondence between the points of the two cameras that correspond to the same coordinates of the projector is obtained. In addition, if a camera observes the patterns from two projectors, the coordinates of the two projectors that correspond to the same point in the camera image are set as the corresponding points.

The extrinsic parameters and focal lengths of cameras and projectors are estimated using bundle adjustment package provided by Lourakis et al. [30]. There remains scale ambiguity after bundle adjustment is applied. The unknown gauge parameters can be decided from real distances between specific cameras.

For the system used for the experiments of the present paper, we used a white screen for projectors as the target object. The front side of the screen was directed to each of the six cameras, and captured. For each of the positions, correspondence information is obtained from three cameras that can capture the front side. Finally, the calibration was processed from 18 ( $3 \times 6$ ) sets of correspondence information.

## 4. Theory and Algorithm

### 4.1 Shape from Multiple Pencil Pattern

In this paper, we propose a shape acquisition method with multiple projectors and multiple cameras. First, we describe the proposed reconstruction method for a minimum configuration of two projectors, a camera for observing the curves and their intersection points (referred to as camera 1 in this section), and an auxiliary camera to solve ambiguity (camera 2 in this section).

The projected parallel lines sweep a set of planes that intersect along a single line. These planes are elements of a pencil of planes, and each can be specified by a single parameter. Using this representation, simple linear constraints can be obtained from the detected line intersections.

Kawasaki et al. [3] used a single projector that projects both vertical and horizontal lines (e.g., a grid pattern). In this configuration, the axes of two pencils of planes (vertical and horizontal) intersect at the optical center of the projector. This enables the elimination of all of the constant terms of the linear equations obtained from the grid points. Thus, the linear equations have 1-DOF indeterminacies. They determined a unique solution using a modulated grid pattern in which the intervals between lines are varied randomly.

Different from the above method, each of the two projectors projects a set of parallel lines. And each set of parallel lines forms a pencil of planes. We assume that the projectors are positioned



so that the axes lines of these pencils of planes are skew (i.e., do not intersect in 3D space). Theoretically, in this configuration, the projected patterns observed by camera 1 can be uniquely solved, as described in the following in this section, with its proof shown in Appendix A.2. However, since this solution often becomes unstable for real applications, we propose a method for solving the instability using camera 2 which is shown in Section 4.2.

In the following discussion, the 3D space is represented in the camera coordinate system of camera 1, in which the origin of 3D space is at the optical center of the camera the  $xy$ -plane is parallel to the image plane, and the direction of  $z$ -axis is toward backward of the camera (a right hand coordinate system with  $z$ -axis pointing outward).

When a line pattern is projected by a projector, it sweeps a 3D plane in the space, and the illuminated points on the target scene is on that plane. We call this plane as a pattern plane. In the proposed method, a pattern plane  $p$  is represented by an equation

$$p_1x + p_2y + p_3z + 1 = \mathbf{p}^T \mathbf{x} + 1 = 0, \quad (1)$$

where vector  $\mathbf{x} = (x, y, z)^T$  is an arbitrary 3D point on pattern plane  $p$ , and vector  $\mathbf{p} = (p_1, p_2, p_3)^T$  is a parameter vector of the plane. Physically,  $\mathbf{p}$  is a normal vector of the pattern plane with its scale representing a reciprocal number of the (signed) distance between the origin and the plane.

Let an axis of vertical planes be  $l_v$ . Then, all of the vertical patterns include  $l_v$ , and  $l_v$  includes the optical center of the projector. Then, the set of planes that includes  $l_v$  can be described by a single parameter  $\mu$  as follows:

$$\mathbf{v}(\mu)^T \mathbf{x} + 1 = 0, \quad \mathbf{v}(\mu) \equiv \mathbf{v}_0 + \mu \mathbf{v}_d, \quad (2)$$

where  $\mathbf{v}(\mu)$  represents the scaled normal vector of a plane in the set of planes, and  $\mathbf{v}_0$  and  $\mathbf{v}_d$  are constant vectors that can be determined from the position of line  $l_v$ . Detailed derivation of Eq. (2) is described in Appendix A.1.

Similarly, the set of horizontal planes that include an axis  $l_h$  is described by the parameter  $\rho$  as follows:

$$\mathbf{h}(\rho)^T \mathbf{x} + 1 = 0, \quad \mathbf{h}(\rho) \equiv \mathbf{h}_0 + \rho \mathbf{h}_d, \quad (3)$$

where  $\mathbf{h}(\rho)$  represents the scaled normal vector of a plane, and  $\mathbf{h}_0$  and  $\mathbf{h}_d$  are constant vectors that represents the position of line  $l_h$ .

Suppose that an intersection between vertical pattern  $v$  and horizontal pattern  $h$  is detected by camera 1, and its position is at  $(s, t)$  in the image coordinates of a normalized camera (image coordinates represented in unit of the focal length with the principle point at the origin). Let the planes of patterns  $v$  and  $h$  be represented by Eqs. (2) and (3), respectively. Then,

$$\mathbf{u}^T (\mathbf{v}(\mu) - \mathbf{h}(\rho)) = 0, \quad (4)$$

where  $\mathbf{u} \equiv (s, t, -1)^T$  is obtained as the constraint from the intersection between planes  $v$  and  $h$  [4], [31]. Using Eqs. (2) and (3), we obtain

$$\begin{aligned} & \mathbf{u}^T (\mathbf{v}_0 + \mu \mathbf{v}_d - \mathbf{h}_0 - \rho \mathbf{h}_d) \\ &= (\mathbf{u}^T \mathbf{v}_d) \mu - (\mathbf{u}^T \mathbf{h}_d) \rho + \mathbf{u}^T (\mathbf{v}_0 - \mathbf{h}_0) = 0, \end{aligned} \quad (5)$$

where  $\mu$  and  $\rho$  are unknown variables and the rest are known constants.

Since Eq. (5) can be obtained for each intersection point, a system of simultaneous linear equations can be obtained from the detected image. Let  $\mu_i$  be the parameter  $\mu$  in Eq. (2) of the plane formed by the  $i$ -th detected vertical pattern. Similarly, let  $\rho_j$  be the parameter  $\rho$  (Eq. (3)) of the  $j$ -th detected horizontal pattern. Suppose that  $K$  intersections are detected. We also define maps  $\alpha$  and  $\beta$  from indices of intersections to indices of detected patterns, such that the  $k$ -th intersection point ( $\mathbf{u}_k = (s_k, t_k, -1)$ ) is an intersection between the  $\alpha(k)$ -th detected vertical pattern and the  $\beta(k)$ -th detected horizontal pattern. Then, we obtain

$$\begin{aligned} A_k &\equiv \mathbf{u}_k^T \mathbf{v}_d, \quad B_k \equiv \mathbf{u}_k^T \mathbf{h}_d, \quad C_k \equiv \mathbf{u}_k^T (\mathbf{v}_0 - \mathbf{h}_0), \\ A_k \mu_{\alpha(k)} - B_k \rho_{\beta(k)} &= -C_k, \end{aligned} \quad (6)$$

for  $k = 1, \dots, K$ .

By solving the simultaneous equations obtained from Eq. (6), the planes formed by the detected patterns can be solved. The equation becomes

$$\mathbf{M} \mathbf{x} = \mathbf{c}, \quad (7)$$

where  $\mathbf{x}$  is a vector that consists of  $\mu_i$  and  $\rho_j$ .

If a sufficient number of intersections are provided, then the number of constraints is greater than the number of variables, and the solution of Eq. (7) is computed by  $\mathbf{x} = (\mathbf{M}^T \mathbf{M})^{-1} \mathbf{M}^T \mathbf{c}$ . The computation of  $(\mathbf{M}^T \mathbf{M})^{-1}$  is performed by  $LU$  decomposition. For the case of four patterns shown in Fig. 4, a proof in which Eq. (7) has a unique solution is provided in Appendix A.2.

By applying triangulation to the planes of the solution, 3D shape reconstruction of the detected patterns can be achieved.

In work of Sagawa et al. [4], Eq. (6) is used for a configuration of one camera and one projector. In this case, the constant term of Eq. (6) becomes zero, and thus, Eq. (7) has 1-DOF indeterminacy (scaling of  $\mathbf{x}$ ). Sagawa et al. solved this indeterminacy by using constraints that each pattern plane of Eqs. (2) and (3) should coincide with a finite set of planes that are known from

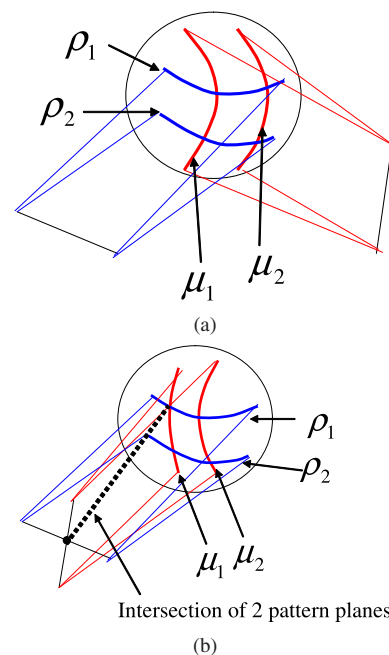


Fig. 4 Positions of four pattern planes: (a) the proposed method, (b) previous methods [3], [4], [31].

calibration. Although this problem can be solved even with a uniformly-spaced grid pattern, they used modulated grid pattern to achieve stable results.

On the contrary, using the method described in this section with two sets of parallel-line patterns projected, the solution of Eq. (7) becomes unique. Thus, theoretically, the solution can be directly obtained by just solving the equation. However, in reality, directly solving Eq. (7) often produces large errors. The reason and solution for this problem will be described in the next section.

#### 4.2 Multi-view Reconstruction with Error Correction

In the previous section, it was shown that, if two projectors are used, theoretically, a unique solution can be obtained from the linear equations. However, in reality, the solution of linear equation (7) can be unstable if the equation is “nearly” degenerated. For example, if the distribution of the direction vectors of  $K$  intersections  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_K$  is in a small area (i.e.,  $\mathbf{u}_1 \approx \mathbf{u}_2 \approx \dots \approx \mathbf{u}_K$ ), then the minimum eigenvalue of  $(\mathbf{M}^T \mathbf{M})^{-1}$  becomes approximately 0, which is nearly a degenerate condition, making the calculation of the linear solution unstable.

In contrast, Eq. (7) can be solved stably up to 1-DOF ambiguity. Intuitively, this is explained as follows. If the parameter of curve 1,  $\mu_1$ , is fixed (i.e., the pattern plane of curve 1 is fixed), then the shape of curve 1 is reconstructed. Using these depth values, the pattern planes of the curves that have intersections with curve 1 can be decided. By repeating these processes, all of the connected curves can be calculated. Thus, Eq. (7) is stable except for the 1-DOF ambiguity.

Theoretically, the proposed method can be used to obtain a unique solution from only linear equations by using a system with two projectors and a camera. However, the linear solution sometimes becomes unstable when the minimum eigenvalue of matrix  $(\mathbf{M}^T \mathbf{M})^{-1}$  becomes approximately zero (i.e., nearly degenerate condition). Assume an extreme condition in which the intersections  $\mathbf{u}_i = (u_i, v_i, -1)$  are all the same. Then, it can easily be confirmed that the system of equations (6) has a 1-DOF indeterminacy (Note that such an error is similar to an indeterminacy mentioned in Refs. [3], [4], [5]). Although this is an extreme case, the situation can be nearly degenerate if the angle of view of the camera is small or if the region of the connected grid has only a small area. Such conditions sometimes occur in actual systems. In this situation, the solution becomes sensitive to errors in calibration and in the line detection step.

In the present paper, we propose a method by which to reduce the 1-DOF error by using the constraint between multiple cameras. When a 3D curve reconstructed from a line by a pair of Camera 1 and Projector 1 is reprojected to Camera 2 of the system, the projection of the curve overlaps the detected line of Camera 2 if there is no occlusion and the 3D reconstruction is correct. In addition, in order to further stabilize the solution and reduce the computational complexity, we use periodic IDs of the de Bruijn sequence obtained in the line detection step. An overview of this method is given in the following.

First, we define a one-parameter representation of the solution, as follows. Let the linear equation be  $\mathbf{M}\mathbf{x} = \mathbf{c}$ , and let the solution be  $\mathbf{x}_0$ . Due to the error in  $\mathbf{M}$  and  $\mathbf{c}$ ,  $\mathbf{x}_0$  has some error. Since

the distribution of the errors can be explained by 1-DOF indeterminacy, let the correct solution  $\bar{\mathbf{x}}$  be  $\bar{\mathbf{x}} = \mathbf{x}_0 + s\mathbf{e}$ , where  $s$  is a parameter and  $\mathbf{e}$  is the direction vector of the space of the error. Then,  $\mathbf{e}$  can be computed as the vector that minimizes  $\|\mathbf{M}\mathbf{e}\|$ , which is the eigenvector corresponding to the minimum eigenvalue of  $\mathbf{M}^T \mathbf{M}$ . By searching the parameter  $s$  using a criterion that is different from the linear equation, the proposed method improves the solution. The actual algorithm becomes as follows:

**Step 1** Iterate the following steps (from 1.1 to 1.3) by changing  $s$  such that the  $\mathbf{x}(s) = \mathbf{x}_0 + s\mathbf{e}$  sweeps the possible positions of the pattern planes.

**Step 1.1** From the candidate solution  $\mathbf{x}(s) = \mathbf{x}_0 + s\mathbf{e}$ , decide the correspondences between the detected curves and the line patterns on the projector, by finding the nearest line pattern that has the same de Bruijn ID to the solution of a detected curve.

**Step 1.2** Re-project the 3D curve computed from the correspondences to the adjacent camera 2.

**Step 1.3** Compare the reprojected curves with the curves detected in Camera 2 by calculating a cost function that represents distances between those curves. The cost function is basically the sum of distances from all the points from the reprojected curves to the detected curves. The distances are computed only if the de Bruijn IDs of the curves are the same.

**Step 2** Select  $s$  of the minimum cost function. Output the correspondences between the detected curves and the line patterns computed for this  $s$ .

The definition of the cost function of Step 1.2 to evaluate  $\mathbf{x}(s)$  is as follows. From  $\mathbf{x}(s)$ , parameter information of pattern planes that illuminate the observed curves can be obtained. Using this information, ID of pattern that corresponds each of the observed curves is decided. Since patterns on the projector image is finite and the pattern plane for each pattern is known from the calibration, we select the nearest pattern for each of the pattern plane parameter obtained from  $\mathbf{x}(s)$ . Then, mapping from the observed curves to the finite set of patterns can be decided. Finally, using this mapping the 3D positions of the observed curves are reconstructed by triangulation, and the resulting 3D curves are projected to camera 2 for evaluation.

Let  $c_i(s)$  be a projection of curve  $i$  reconstructed from  $\mathbf{x}(s)$  to camera 2. We define the matching score of  $\mathbf{x}(s)$  by:

$$S_T(s) \equiv \sum_i S_C(c_i(s), K), \quad (8)$$

where  $K$  is the set of 2D curves,  $S_C(c, K)$  is a matching score between a 2D curve  $c$  and  $K$ ,  $S_T(s)$  is the total matching score for all the curves reconstructed from  $\mathbf{x}(s)$ . The score becomes small if the curves match better.

About the definition of  $S_C(c, K)$ . Many of the curves observed by camera 1 cannot be observed by camera 2, because there are occlusions or detection failures. Thus, if  $S_C(c, K)$  is defined to be a plus value when a corresponding curve of curve  $c$  cannot be found in  $K$ , the plus values of unobserved curves are accumulated to the cost function. This may lead to incorrect evaluation of matching.

To deal with this problem,  $S_C(c, K)$  is defined as 0 if a corresponding curve of curve  $c$  cannot be found in  $K$ , otherwise, it is defined such that  $S_C(c, K) < 0$ , where

$$S_C(c, K) \equiv \sum_{x_c \in c} \min(0, \min_{k \in K} \min_{x_k \in k} (\|x_c - x_k\| - W)), \quad (9)$$

where  $x_c$  is a point on  $c$ ,  $k$  is a curve in  $K$ ,  $x_k$  is a point on  $k$ , and  $W$  is a size of neighborhood from  $c$ . Points within distance  $W$  from curve  $c$  are seen as nearby points to the curve, and affect the score. Points that are not within distance  $W$  do not affect the score because of the first min function.

### 4.3 Detecting Lines from Multiple Projectors

The method for line detection is based on work of Sagawa et al. [4]. The number of line directions in their paper was only two, i.e., vertical and horizontal. However, in the case considered in the present study, since the entire shape is projected by multiple projectors, patterns of more than three directions are observed in a captured image. Therefore, it is necessary to discriminate the patterns projected from different projectors. In addition, the method in Sagawa's work [4] detects nearly vertical lines in order to discriminate them from horizontal lines. However, if more than three patterns are projected, the relative angle between different patterns is smaller than 90 degrees. This causes misclassification of lines.

We therefore propose two methods by which to discriminate lines. First, we used different colors for the vertical (or horizontal) patterns that can be observed simultaneously from a camera. In the present paper, two color pairs are used: blue and cyan, and red and yellow. By changing the colors used for the color code, the detection of lines from different projectors can be suppressed. Since curve detection algorithm of Sagawa et al. uses only two colors for classification and curve detection itself is processed on single channel image, their detection method is stable to effects of textures on the target object (a textured plane of a book is reconstructed in their work [4]). Since the curve detection method of this paper is based on their work, although the total number of colors are increased to four, the proposed method is also stable to textures.

The color component used for curve detection are blue or red, which depends on the color pair used. Since the crosstalk between blue and red channels is negligible, the patterns casted from different projectors can be discriminated. The color identification between blue and cyan, or red and yellow is done by using green channel of detected curves. The proposed method identifies the colors by thresholding the ratio of green/blue or green/red. The threshold is determined by manual in the current implementation.

The second approach is to rotate the image to be processed. For example, in Fig. 5, the patterns are projected from three projectors: one vertical pattern and two tilted patterns. If the vertical pattern is to be detected, one of the tilted patterns can be detected because the color is still the same. Therefore, we slightly rotate the image in the counter-clockwise direction in order to detect the vertical pattern and suppress detection of the tilted pattern. Before the post-processing, including color decoding and 3D re-

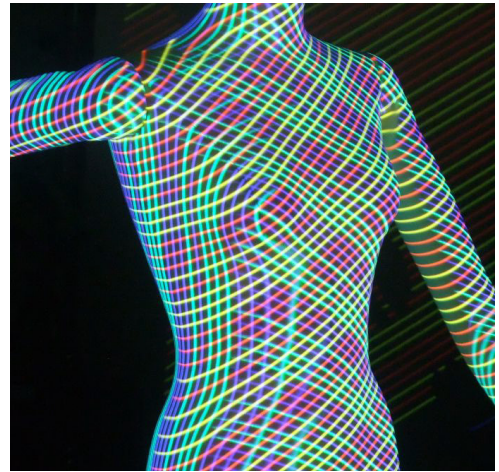


Fig. 5 Line patterns projected from three projectors.

construction, the positions of detected lines are rotated back to the original positions.

### 4.4 Filtering Erroneously Reconstructed Curves

Using the method described above, the correspondences from the detected curves to the patterns can be determined. These correspondences may include errors caused by various reasons such as incorrect connection between different curves or detection errors of color codes.

Typically, in this case, curves with incorrect correspondences are reconstructed at totally different positions from the correct position. At a detected intersection on the incorrect curve, there are normally 3D gaps between the 3D curves reconstructed from the two curves that form the intersection (Note that there are no gaps between the 3D curves at the intersection if the intersecting curves are correctly reconstructed).

In order to filter out such curves, the intersection points used for reconstruction are re-checked for the reconstructed curves. If the distance between the 3D points on the 3D curves exceeds a tolerance threshold, then the points are labeled as invalid. Curves that have numerous invalid intersection points are rejected from the reconstruction.

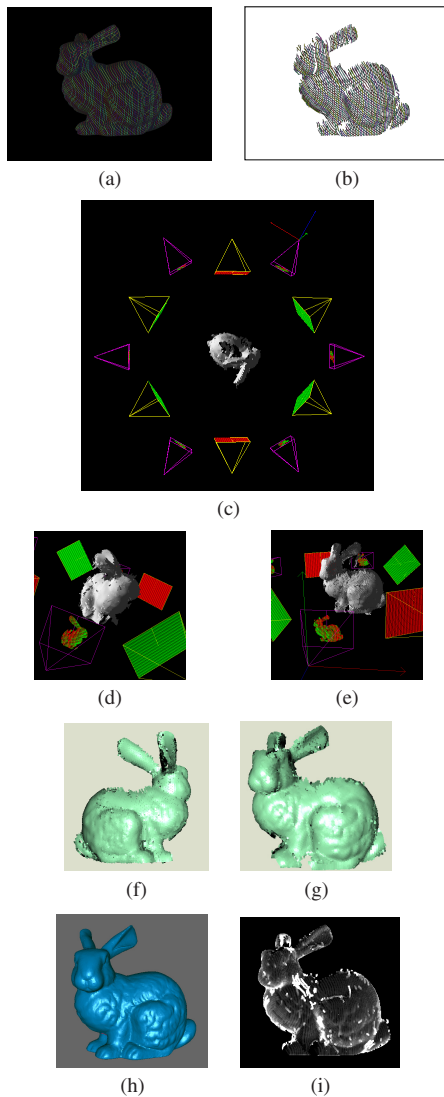
## 5. Experiments

### 5.1 Evaluation of Entire Shape Acquisition Using Synthetic Data

In order to confirm the effectiveness of the proposed method, synthetic data is used for reconstruction. We use the Stanford bunny as the target. We virtually set six cameras and six projectors surrounding the target object and render the captured images using the POV-Ray. The synthesized images are shown in Fig. 6 (a). We apply the curve decomposition and the detection method to the images, and the detection results are shown in Fig. 6 (b). Based on these detection results, we can observe that lines near occluding boundaries were compressed and were not detected, whereas most of the curves were successfully detected.

Using the intersection of detected curves of two adjacent projectors, simultaneous equations are constructed and shapes are reconstructed by solving the equations with the proposed multi-view reconstruction technique as shown in Fig. 6 (c)–(g). Fig-



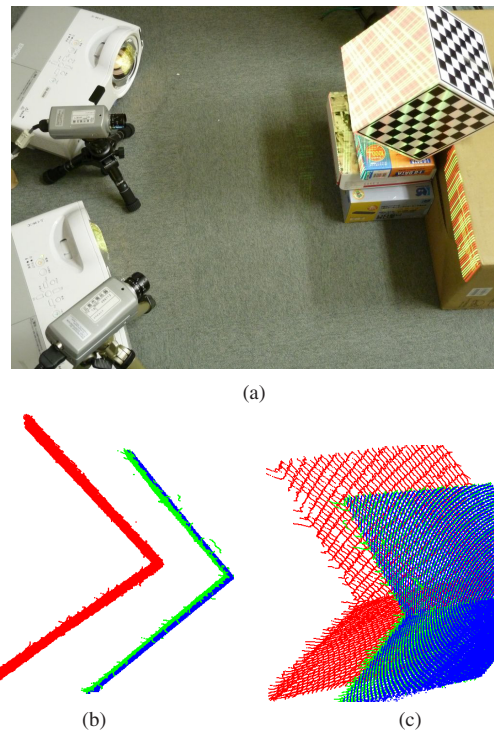


**Fig. 6** Reconstruction results: (a) Image synthesized by POV-Ray (b) Detected curves with de Bruijn ID. (c) through (e) Recovered shape rendered with positions of the cameras and the projectors. (f) and (g) Shaded results. (h) Ground truth. (i) Differences between ground truth and the result of the proposed method.

ure 6(h) shows the ground truth. Average error and RMSE of reconstructed shapes from ground truth are 0.0042 and 0.0089, respectively (the height of the bunny was scaled to be 0.1). Figure 6(i) shows the difference map between ground truth and our result (the maximum intensity value  $R = 255, G = 255, B = 255$  is set to be error value 0.01). These figures indicate that most of the parts were successfully recovered, whereas the true shapes of several small parts (curves) and parts near occluding boundaries were not recovered. Those wrongly reconstructed curves are mainly caused by curve detection errors and the solution is our future research.

### 5.2 Evaluation of Single-side Acquisition Using a Real System

Next, we constructed an experimental system in order to evaluate the multi-view reconstruction method for projector-camera system described in Section 4.2 using two cameras (Sony SX-910s) and two LCD video projectors having SXGA resolution, as shown in Fig. 7(a). Using the experimental system, we measured



**Fig. 7** Evaluation of multi-view reconstruction: (a) Experimental system composed of two cameras and two projectors. (b) and (c) Reconstruction results. Red, green, and blue points are the results obtained by method A, method B, and Gray codes (ground truth up to calibration errors).

**Table 1** Evaluation of accuracy improvement using multi-view reconstruction for GOS.

Evaluation values	Method A	Method B
Angle between 2 faces (deg.)	97.7	90.1
RMSE of points from fit planes	0.00094	0.00181

a cube-shaped object using two methods. In method A, 3D reconstruction is performed using only linear equations (6) using two projectors and a single camera. In method B, 3D reconstruction is further improved using the second camera, using the multi-view reconstruction algorithm. As the ground truth, we also measured the target object using Gray codes, in which the correspondences are correct (i.e., the 3D reconstruction is correct up to calibration errors).

The 3D reconstruction results are shown in Fig. 7(b) and 7(c). In order to evaluate the reconstruction results of methods A and B, two faces of the cube-shaped object are extracted and fitted to 3D planes, and the angles between the fitted planes (which should be right angles) and the RMSEs from the fitted planes were measured, as shown in Table 1. Since the true scale was not known, the RMSEs were normalized, such that the average distances between the points and the camera became 1.0.

As shown in Fig. 7(b) and 7(c), the results of method A (red) include errors from the ground truth (blue), whereas the results of method B (green) approximately coincide with the ground truth. (The bias from the ground truth is caused by the curve detection algorithm.) About the angle between the faces in Table 1, the ground truth is 90 (the right angle). Thus, the result of the method B is much more precise than method A for errors of the measured angle.

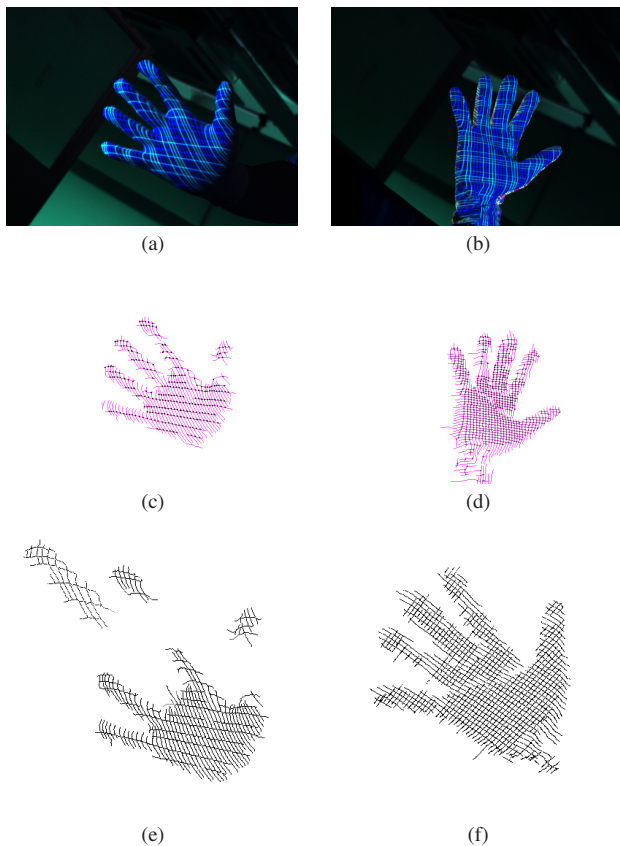
About the RMSEs from the fit planes in Table 1, the result of



method A is smaller than method B. This is because, in method A, the distances between the horizontal and vertical patterns at the intersection points are minimized at the expense of the correctness of the global shape. This is natural, since method A is based only on linear equations (6), which reduces distances of the intersecting curves at the intersection points (thus, reduces RMSEs of fit planes). In method B, the observed curves are corresponded to the finite set of “known” planes using information from multiple cameras, which increase the RMSEs because there are always unavoidable calibration errors.

**5.3 Comparison of the Proposed Method and the One-shot Scanning Method Using One Projector**

We compare the proposed method with the one-shot scanning method using one projector [4] (referred to herein as the one-projector method). The proposed method can use a more dense pattern than the one-projector method because the 1D search for shape reconstruction is not necessary. The advantage of the proposed method is that reconstruction of small objects, such as thin fingers, is easier than in the one-projector method. **Figure 8** shows the results of reconstructing a hand for both of these methods. The top row shows the input images. The middle row shows the grid graph obtained by line detection. The bottom row shows the reconstructed shape. The proposed method (right column) successfully reconstructed all of the fingers, whereas the one-projector method (left column) was not able to achieve this due to the insufficient density of the grid graph.



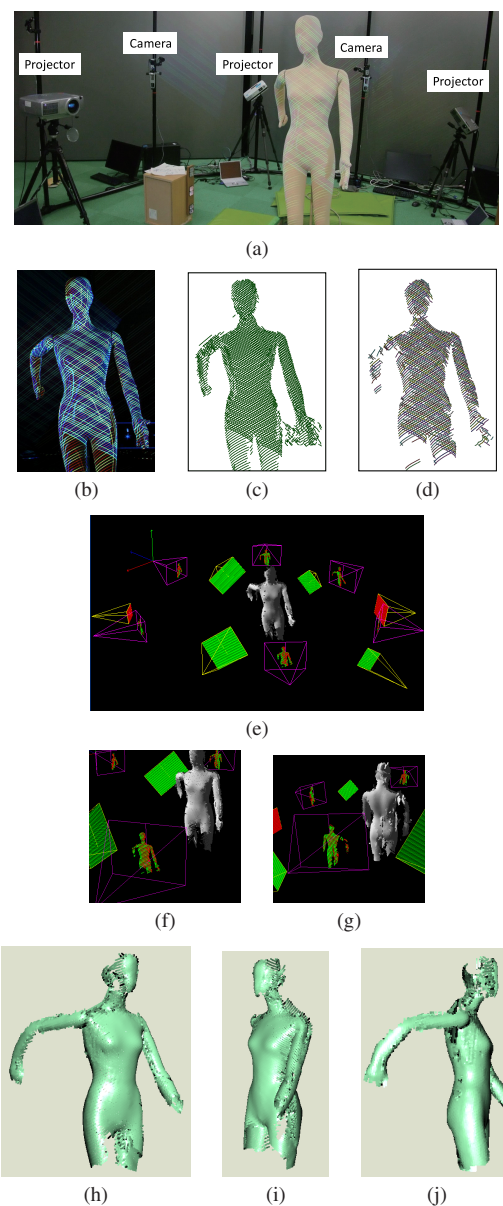
**Fig. 8** Hand reconstructed by Ref. [4] (left) and the proposed method (right). The top row shows the input images. The middle row shows the grid graph obtained by line detection. The bottom row shows the reconstructed shape.

**5.4 Evaluation of Entire Shape Acquisition Using a Real System**

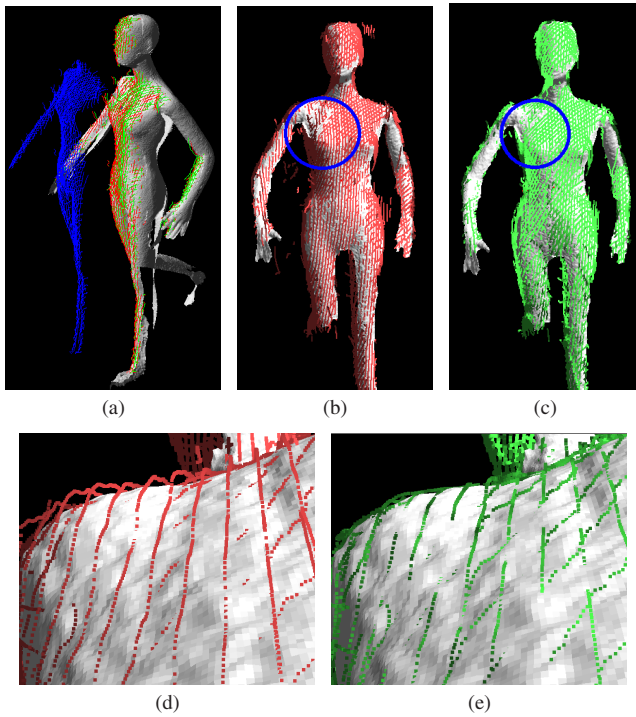
Finally, we constructed an experimental system with which to capture an entire shape with six cameras and six projectors, as shown in **Fig. 9** (a). The cameras are Point Grey Research Flea2 (SXGA resolution) cameras, and the projectors are LCD video projectors of SXGA resolution. As a target object, we measured a mannequin of 1.7 m in height.

One of the captured images is shown in **Fig. 9** (b). We apply the curve decomposition and detection method to the images, and the detection results are shown in **Fig. 9** (c) and 9 (d).

Using the intersection of the detected curves of two adjacent projectors, the shapes are reconstructed by solving the equations using the multi-view reconstruction method shown in **Fig. 9** (e) through 9 (j). These figures indicate that the large areas are suc-



**Fig. 9** Experiments using a real system: (a) A configuration of six projectors and six cameras. This figure shows a part of the system, namely, three projectors and two cameras. (b) Image captured by one camera. (c) Detected lines. (d) de Bruijn ID. (e) through (g) Recovered shape rendered with positions of the cameras and the projectors. (h) Shaded results.



**Fig. 10** Results of multi-view reconstruction: (a) Blue: result without error correction using multi-view information (result A), red: result of multi-view reconstruction (result B), green: result after filtering erroneous curves (result C). (The shaded model is a ground-truth polygon reconstructed by a Gray code.) (b) and (c) Results of B and C from different viewpoints. (d) and (e) Close-up views of (b) and (c), respectively.

cessfully recovered, whereas several parts that are not observed by cameras nor projected by the projector were not reconstructed. Other parts that were not recovered were small areas that were disconnected from larger parts. This is because the system of equations with small areas having a small number of lines tends to become a system of nearly degenerate equations.

For evaluation, we also measured the mannequin without applying the proposed methods. First, 3D reconstruction was performed using only linear equations [32] for each projector-camera system, which consists of a camera and two projectors (result A). The error correction technique using multi-view information described in Section 4.2 was then applied (result B). Finally, 3D reconstruction was further improved by noise filtering (result C), as described in Section 4.4. As the ground truth, we also measured the object using a Gray code pattern, in which the correspondences are correct (i.e., the 3D reconstruction is correct up to calibration errors).

The 3D reconstruction results of A, B, and C are shown in Fig. 10(a) through 10(c), and close-up views are shown in Fig. 10(d) and 10(f). In order to evaluate the reconstruction results, the RMSEs were calculated, as shown in Table 2.

As shown in the figures, the results obtained without using multi-view information (result A, shown as blue points) include significant errors (large gaps because of incorrect 1-DOF estimation) from the ground truth (shaded polygon) in Fig. 10(a), whereas the final results of the multi-view reconstruction (results B and C, shown as green points) approximately coincide with the ground truth (Fig. 10(c) and 10(e)). From Table 2, we can

**Table 2** RMSE (mm) values for each method.

A	B	C
Linear solution [32]	multi-view reconstruction	B + Filtering
291.4	9.0	7.6

also confirm that the results obtained by the proposed method are significantly improved by the proposed multi-view reconstruction method. The main reason for the large errors of result A is that, in the method used for A [32], a special arrangement of the projectors and cameras is required in order to achieve a stable result, whereas, in the experiment of the present study, we freely set up the projectors and the cameras.

## 6. Conclusion

In the present paper, an active one-shot scanning system using multiple projectors and cameras to capture the entire shape of an object was presented. One critical problem in using multiple projectors simultaneously is that multiple patterns projected onto the same object interfere with each other and it is difficult to decompose these patterns. In order to address this problem, we proposed a new method in which each projector projects only a set of parallel lines with two colors. Reconstruction is achieved by solving linear equations derived from intersection points between lines. We also implemented a multi-view reconstruction technique for the proposed projector-camera system in order to improve accuracy. In the experiment of the present study, we constructed a system that consists of six projectors and six cameras and captured moving object successfully. Scanning of the entire shape of an object exhibiting dynamic motion is an important extension of the proposed method. In the future, we intend to develop a system that uses infrared projectors that does not interfere with the visible appearance of the scenes. Using infrared light patterns, we intend to capture dynamic scenes and visible textures simultaneously.

**Acknowledgments** This work was supported in part by Strategic Information and Communications R&D Promotion Programme (SCOPE) No.101710002 (Ministry of Internal Affairs and Communications, Japan), Grant-in-Aid for Scientific Research No.21200002 (Ministry of Education, Science, Sports and Culture, Japan) and Funding Program for Next Generation World-Leading Researchers No.LR030 (Cabinet Office, Government of Japan) in Japan.

## Reference

- [1] Laurentini, A.: How far 3d shapes can be understood from 2d silhouettes, *IEEE Trans. PAMI*, Vol.17, No.2, pp.188–195 (1995).
- [2] Salvi, J., Pages, J. and Batlle, J.: Pattern Codification Strategies in Structured Light Systems, *Pattern Recogn.*, Vol.37, No.4, pp.827–849 (2004).
- [3] Kawasaki, H., Furukawa, R., Sagawa, R. and Yagi, Y.: Dynamic scene shape reconstruction using a single structured light pattern, *CVPR 2008*, pp.1–8 (2008).
- [4] Sagawa, R., Ota, Y., Yagi, Y., Furukawa, R., Asada, N. and Kawasaki, H.: Dense 3D reconstruction method using a single pattern for fast moving object, *ICCV 2009*, pp.1779–1786 (2009).
- [5] Ulusoy, A.O., Calakli, F. and Taubin, G.: One-Shot Scanning using De Bruijn Spaced Grids, *The 7th IEEE Conf. 3DIM*, pp.1786–1792 (2009).
- [6] Furukawa, Y. and Ponce, J.: Accurate, Dense, and Robust Multi-View Stereopsis, *IEEE Trans. PAMI*, Vol.32, No.8, pp.1362–1376 (2010).
- [7] Furukawa, Y. and Ponce, J.: Dense 3D Motion Capture from Synchronized Video Streams, *CVPR 2008*, pp.1–8 (2008).

- [8] Seitz, S.M., Curless, B., Diebel, J., Scharstein, D. and Szeliski, R.: A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms, *CVPR 2006*, pp.519–528 (2006).
- [9] Scholz, V., Stich, T., Keckeisen, M., Wacker, M. and Magnor, M.: Garment Motion Capture Using Color-Coded Patterns, *Computer Graphics Forum (Proc. Eurographics EG'05)*, Vol.24, No.3, pp.439–448 (2005).
- [10] White, R., Crane, K. and Forsyth, D.: Capturing and Animating Occluded Cloth, *ACM Trans. Gr. (SIGGRAPH)*, Vol.26, No.3 (2007).
- [11] Bradley, D., Popa, T., Sheffer, A., Heidrich, W. and Boubekeur, T.: Markerless Garment Capture, *ACM Trans. Gr. (SIGGRAPH)*, Vol.27, No.3, p.99 (2008).
- [12] Kutulakos, K.N. and Seitz, S.M.: A Theory of Shape by Space Carving, *IJCV*, Vol.38, pp.307–314 (2000).
- [13] Franco, J.-S. and Boyer, E.: Fusion of Multi-View Silhouette Cues Using a Space Occupancy Grid, *ICCV 2005*, Vol.2, pp.1747–1753 (2005).
- [14] Kinoca Minolta Sensing Americas, Inc.: Vivid 910 non-contact digitizer, available from (<http://sensing.konicaminolta.us/>).
- [15] Leica Geosystems, available from (<http://www.leica-geosystems.com/>).
- [16] PMD Technologies, available from (<http://www.pmdtec.com/>).
- [17] Mesa Imaging AG.: SwissRanger SR-4000, available from (<http://www.swissranger.ch/index.php>).
- [18] Young, M., Beeson, E., Davis, J., Rusinkiewicz, S. and Ramamoorthi, R.: Viewpoint-Coded Structured Light, *CVPR 2007*, pp.1–8 (2007).
- [19] Battle, J., Mouaddib, E.M. and Salvi, J.: Recent progress in coded structured light as a technique to solve the correspondence problem: A survey, *Pattern Recogn.*, Vol.31, No.7, pp.963–982 (1998).
- [20] Lanman, D., Crispell, D. and Taubin, G.: Surround Structured Lighting for Full Object Scanning, *3DIM 2007*, pp.107–116 (2007).
- [21] Je, C., Lee, S.W. and Park, R.-H.: High-Contrast Color-Stripe Pattern for Rapid Structured-Light Range Imaging, *ECCV*, Vol.1, pp.95–107 (2004).
- [22] Vuylsteke, P. and Oosterlinck, A.: Range Image Acquisition with a Single Binary-Encoded Light Pattern, *IEEE Trans. PAMI*, Vol.12, No.2, pp.148–164 (1990).
- [23] Zhang, L., Snavely, N., Curless, B. and Seitz, S.M.: Spacetime Faces: High-Resolution Capture for Modeling and Animation, *ACM Annual Conference on Computer Graphics*, pp.548–558 (2004).
- [24] Davis, J., Nehab, D., Ramamoorthi, R. and Rusinkiewicz, S.: Spacetime Stereo: A Unifying Framework for Depth from Triangulation, *IEEE Trans. PAMI*, Vol.27, No.2, pp.296–302 (2005).
- [25] Furukawa, R. and Kawasaki, H.: Uncalibrated multiple image stereo system with arbitrarily movable camera and projector for wide range scanning, *The 5th IEEE Conf. 3DIM*, pp.302–309, IEEE (2005).
- [26] Salvi, J., Battle, J. and Mouaddib, E.M.: A robust-coded pattern projection for dynamic 3D scene measurement, *Pattern Recogn.*, Vol.19, No.11, pp.1055–1065 (1998).
- [27] Zhang, L., Curless, B. and Seitz, S.M.: Rapid Shape Acquisition Using Color Structured Light and Multi-Pass Dynamic Programming, *3DPVT*, pp.24–36 (2002).
- [28] Bradski, G. and Kaehler, A.: *Learning OpenCV: Computer Vision with the OpenCV Library*, O'Reilly Media (2008).
- [29] Zhang, S. and Huang, P.S.: Novel Method for Structured Light System Calibration, *Optical Engineering*, Vol.45, No.8, pp.083601 (2006).
- [30] Lourakis, M.A. and Argyros, A.: SBA: A Software Package for Generic Sparse Bundle Adjustment, *ACM Trans. Math. Software*, Vol.36, No.1, pp.1–30, DOI: 10.1145/1486525.1486527 (2009).
- [31] Furukawa, R., Kawasaki, H., Sagawa, R. and Yagi, Y.: Shape from grid pattern based on coplanarity constraints for one-shot scanning, *IPSJ Trans. Computer Vision and Applications*, Vol.1, pp.139–157 (2009).
- [32] Kawasaki, H., Furukawa, R., Sagawa, R., Ohta, Y., Sakashita, K., Zushi, R., Yagi, Y. and Asada, N.: Linear solution for oneshot active 3D reconstruction using two projectors, *3DPVT 2010*, available from (<http://campwww.informatik.tu-muenchen.de/3DPVT2010/data/media/e-proceeding/index.html>).

## Appendix

### A.1 One-parameter Representation of Pattern Planes

In this appendix section, detailed derivation of Eq. (2), one-parameter representation of pattern planes, is described.

Let line  $l_v$  be the axis of the vertical pattern planes (i.e., all the vertical pattern planes include  $l_v$  in 3D space). Let point  $\mathbf{b}_v$  be

an arbitrary point on the optical axis of the projector that projects the vertical patterns, and let vector  $\mathbf{d}_v$  be a unit directional vector of vertical line patterns on the image plane of the same projector. All of the lines and points are expressed in the camera coordinates. Then,  $l_v$  includes  $\mathbf{b}_v$  with a directional vector of  $\mathbf{d}_v$ .

Let vector  $\mathbf{v}$  be a normal vector of a pattern plane  $v$ . Since  $v$  includes  $l_v$ ,  $\mathbf{v} \perp \mathbf{d}_v$ . Thus,

$$\mathbf{v}^\top \mathbf{d}_v = 0. \quad (\text{A.1})$$

Since  $v$  includes  $\mathbf{b}_v$ , by using Eq. (2),  $\mathbf{v}$  satisfies

$$\mathbf{v}^\top \mathbf{b}_v + 1 = 0. \quad (\text{A.2})$$

By regarding Eqs. (A.1) and (A.2) as linear equations with respect to 3D vector  $\mathbf{v}$ , solutions of  $\mathbf{v}$  form a 1D vector space. Thus,  $\mathbf{v}$  can be expressed as Eq. (2).

Since Eqs. (A.1), (A.2), and (2) hold for arbitrary values of  $\mu$ ,  $\mathbf{v}_0^\top \mathbf{d}_v = 0$ ,  $\mathbf{v}_0^\top \mathbf{b}_v = -1$ ,  $\mathbf{v}_d^\top \mathbf{d}_v = 0$ ,  $\mathbf{v}_d^\top \mathbf{b}_v = 0$ .

Vector  $\mathbf{v}_d$  that satisfies  $\mathbf{v}_d^\top \mathbf{d}_v = 0$ ,  $\mathbf{v}_d^\top \mathbf{b}_v = 0$  can be calculated by  $\mathbf{v}_d \equiv \mathbf{d}_v \times \mathbf{b}_v$ .

Vector  $\mathbf{v}_0$  that satisfies  $\mathbf{v}_0^\top \mathbf{d}_v = 0$ ,  $\mathbf{v}_0^\top \mathbf{b}_v = -1$  can be calculated by

$$\mathbf{v}_0 = -\frac{\mathbf{b}_v - (\mathbf{b}_v^\top \mathbf{d}_v) \mathbf{d}_v}{\|\mathbf{b}_v\|^2 - (\mathbf{b}_v^\top \mathbf{d}_v)^2}, \quad (\text{A.3})$$

by using  $\|\mathbf{d}_v\| = 1$ .

This one-parameter representation is also shown presented in previous studies [4], [31].

### A.2 Proof of Uniqueness of Solution When Two Projectors are Used.

In this section, we present a proof that linear equation (7) has a unique solution when two projectors are used with general conditions for four pattern planes shown in Fig. 4 (a) (some special conditions are excluded). In this proof, it will be shown that linear equation (7) is not identically equal to 0.

In the following proof, reductio ad absurdum is used. First, a discussion with no contradiction will be shown when one projector is used [3], [4], [31] and four patterns are projected, as shown in Fig. 4 (b), under the assumption that the determinant of the coefficient matrix is identically equal to zero. Then, it will be shown that, when two projectors are used, as shown in Fig. 4 (a), the same discussion leads to a contradiction under the same assumption.

For the case of one projector [3], [4], [31], linear equation (7) can be  $\mathbf{M}\mathbf{x} = \mathbf{0}$  under some parametrization. Since the true solution  $\mathbf{x}$  is not zero,  $\mathbf{M}\mathbf{x} = \mathbf{0}$  is degenerated, and the determinant of  $\mathbf{M}$  is identically zero. For the case of Fig. 4 (b), Eq. (7) becomes

$$\begin{pmatrix} A_{1,1} & 0 & -B_{1,1} & 0 \\ A_{1,2} & 0 & 0 & -B_{1,2} \\ 0 & A_{2,1} & -B_{2,1} & 0 \\ 0 & A_{2,2} & 0 & -B_{2,2} \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \rho_1 \\ \rho_2 \end{pmatrix} = \begin{pmatrix} C_{1,1} \\ C_{1,2} \\ C_{2,1} \\ C_{2,2} \end{pmatrix}, \quad (\text{A.4})$$

where subscripted indices of  $A_{1,2}$  indicate that the coefficient is defined by Eq. (6) using the intersection point between the vertical pattern plane with parameter  $\mu_1$  and the horizontal pattern plane with parameter  $\rho_2$ . Then, determinant  $|\mathbf{M}|$  of coefficient matrix  $\mathbf{M}$



can be calculated as follows:

$$\begin{aligned}
 |\mathbf{M}| &= -A_{1,1}A_{2,2}B_{1,2}B_{2,1} + A_{1,2}A_{2,1}B_{1,1}B_{2,2} \\
 &= -(\mathbf{u}_{1,1}^\top \mathbf{v}_d)(\mathbf{u}_{2,2}^\top \mathbf{v}_d)(\mathbf{u}_{1,2}^\top \mathbf{h}_d)(\mathbf{u}_{2,1}^\top \mathbf{h}_d) \\
 &\quad + (\mathbf{u}_{1,1}^\top \mathbf{h}_d)(\mathbf{u}_{2,2}^\top \mathbf{h}_d)(\mathbf{u}_{1,2}^\top \mathbf{v}_d)(\mathbf{u}_{2,1}^\top \mathbf{v}_d), \quad (\text{A.5})
 \end{aligned}$$

where  $\mathbf{u}_{i,j}$  is a directional vector from the optical center of the camera to the intersection between the vertical pattern plane with parameter  $\mu_i$  and the horizontal pattern plane with parameter  $\rho_j$ , which can be obtained from the image of the intersection.

In order to evaluate the value of the determinant above, it should be noted that the directional vectors of the intersections  $\mathbf{u}_{1,1}$ ,  $\mathbf{u}_{1,2}$ ,  $\mathbf{u}_{2,1}$ ,  $\mathbf{u}_{2,2}$  are not independent of each other. The 2D positions of the intersections in the image plane of the camera are dependent on the positions of the pattern planes and the depths from the camera to these intersections. Thus, there are constraints between the above vectors for the physically possible configurations.

Suppose that  $\bar{\mu}_1$ ,  $\bar{\mu}_2$ ,  $\bar{\rho}_1$ , and  $\bar{\rho}_2$  are true solutions. Then, for example, the intersection with parameter  $\mathbf{u}_{1,1}$  should satisfy

$$(\mathbf{u}_{1,1}^\top \mathbf{v}_d)\bar{\mu}_1 - (\mathbf{u}_{1,1}^\top \mathbf{h}_d)\bar{\rho}_1 + \mathbf{u}_{1,1}^\top (\mathbf{v}_0 - \mathbf{h}_0) = 0 \quad (\text{A.6})$$

from Eq. (5). Conversely, if  $\mathbf{u}_{1,1}$  satisfies Eq. (A.6),  $\mathbf{u}_{1,1}$  can be varied while maintaining the consistency of the entire solution. This is physically equivalent to moving the intersection points between the pattern planes of  $\mu_1$  and  $\rho_1$  along the intersection line between the planes (dotted line in Fig. 4(b)), where the 3D position of the intersection point is changed without changing positions of the pattern planes. Although the 2D position of the intersection is changed, the solution remains consistent. Thus, determinant  $|\mathbf{M}|$  remains to equal to zero.

For the case of one projector, using the parametrization such that  $\mathbf{v}_0 = \mathbf{h}_0$  holds, Eq. (A.6) becomes equivalent to

$$(\mathbf{u}_{1,1}^\top \mathbf{v}_d)\bar{\mu}_1 = (\mathbf{u}_{1,1}^\top \mathbf{h}_d)\bar{\rho}_1. \quad (\text{A.7})$$

From equation of determinant (A.5) equals zero, we obtain the following equation:

$$\frac{(\mathbf{u}_{1,1}^\top \mathbf{v}_d)}{(\mathbf{u}_{1,1}^\top \mathbf{h}_d)} = (\text{A value independent on } \mathbf{u}_{1,1}). \quad (\text{A.8})$$

The same equation can be obtained from Eq. (A.7). There are no contradictions in this case.

In contrast, in the case of two projectors, the same line of reasoning leads to a contradiction. Let us assume that determinant (A.5) is identically equal to zero, whereas  $\mathbf{u}_{1,1}$  is changed such that Eq. (A.6) is satisfied. In the case of two projectors, Eq. (A.6) cannot be transformed to Eq. (A.7). Thus, in general, Eq. (A.6) yields  $\mathbf{u}_{1,1}^\top (\mathbf{v}_0 - \mathbf{h}_0) \neq 0$ . This means that Eq. (A.6) contradicts Eq. (A.8).

The above contradiction means that, in the case of two projectors, determinant (A.5) is not identically equal to zero. Thus, for the case of two projectors, there should generally be a unique solution.



**Ryo Furukawa** is a lecturer of Faculty of Information Sciences, Hiroshima City University, Hiroshima, Japan. He received his M.E. and Ph.D. from Nara Institute of Science and Technology in 1993 and 1995 respectively. He became a research associate at Hiroshima City University in 1995, a lecturer at the same Uni-

versity in 2007. His research area includes shape-capturing, 3D modeling, appearance sampling and image-based rendering. He has received several awards, including ACCV2007 Songde Ma Outstanding Paper Award (one of three best papers) and PSIVT2010 Best Paper Award. He is a member of IPSJ, IEICE, and IEEE.



**Ryusuke Sagawa** is a research scientist at Service Robotics Research Group, Intelligent Systems Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Japan. He received his B.E. in Information Science from Kyoto University in 1998. He received his M.E. in Information Engi-

neering in 2000 and Ph.D. in Information and Communication Engineering from the University of Tokyo in 2003. He was an assistant professor at the Institute of Scientific and Industrial Research, Osaka University. He stayed at ETH Zurich as a visiting researcher in 2008 and moved to AIST in 2010. His primary research interests are computer vision, computer graphics and robotics (mainly geometrical modeling and visualization).



**Hiroshi Kawasaki** is a professor of Department of Information and Biomedical Engineering at Kagoshima University, Japan. He received his M.E. on Information Engineering in 2000 and Ph.D. degree on Information and Communication Engineering in 2003 from the University of Tokyo, Japan, respectively. He started

working at Saitama University, Japan, in 2003. Prior to Saitama University, he worked at Microsoft Research Redmond, WA, USA in 2000. He also stayed one year at INRIA Rhone-alpes, France, as a visiting professor in 2009. His current research topic is on capturing shapes and textures of moving object. He also interested in photo-realistic and realtime rendering for VR/MR systems and ITS. He published over 100 research papers including ICCV, CVPR, IJCV, EG, MVA and ITSJ in computer vision, graphics and their applications. He also won several awards including Songde Ma Outstanding Paper Award (best paper for ACCV) in 2007, best paper award on PSIVT in 2010 and Nagao Prize (best paper on MIRU) in 2011. He is a member of IPSJ, IEICE, and IEEE.





**Kazuhiro Sakashita** is a doctoral student of the Institute of Scientific and Industrial Research, Osaka University, Osaka, Japan. He received his B.E. from Okayama University in 2007. He received his M.E. from Osaka University in 2011. He became a doctoral student at the same University in 2011. His research area in-

cludes shape-capturing, computer vision, face tracking.



**Yasushi Yagi** is a professor at the Institute of Scientific and Industrial Research, Osaka University. He received his Ph.D. degree from Osaka University in 1991. After working at the Product Development Laboratory, Mitsubishi Electric Corporation, he joined Osaka University in 1990. The international conferences for

which he served as the program/general chair include: RO-BIO2006 (PC), ACCV2007 (PC), ACCV2009 (GC), ACPR2011 (PC) and ACPR2013 (GC). He was an editor of IEEE ICRA CEB (2008–2011). He is an associate editor-in-chief of IPSJ Transactions on CVA. He has received several awards, including ACM VRST2003 Honorable Mention Award and PSIVT2010 Best Paper Award. He is a fellow of IPSJ and a member of IEICE, RSJ, and IEEE.



**Naoki Asada** received his B.Eng., M.Eng. and D.Eng. degrees in Electrical Engineering from Kyoto University, Japan, in 1979, 1981 and 1987, respectively. He was a research associate at Fukui Medical School and Kyoto University, an associate professor at Okayama University, and became a professor in

Faculty of Information Sciences at Hiroshima City University in 1995. He has been the President of Hiroshima City University since 2006. His research interests include computer vision, computer graphics and medical image understanding. He won the Outstanding Paper Award of IPSJ in 1993. He is a member of IPSJ, IEICE in Japan, the Japanese Society for Artificial Intelligence, and the Japanese Society of Medical Imaging Technology.

(Communicated by *Yasuyuki Matsushita*)