

Regular Paper

Shape from grid pattern based on coplanarity constraints for one-shot scanning

RYO FURUKAWA ^{†1} HIROSHI KAWASAKI ^{†2}
RYUSUKE SAGAWA ^{†3} and YASUSHI YAGI ^{†3}

Shape acquisition of moving deformable objects with little texture is important for applications such as motion capture of human facial expression. Several techniques using structured light have been proposed. These techniques can be largely categorized into two main types. The first type temporally encodes positional information of a projector's pixels using multiple projected patterns, and the second spatially encodes positional information into areas or color spaces. Although the former technique allows dense reconstruction with a sufficient number of patterns, it has difficulty in scanning objects in rapid motion. The latter technique uses only a single pattern, so it is more suitable for capturing dynamic scenes; however, it often uses complex patterns or color intensities, which are susceptible to noise, pattern discontinuity caused by edges, or textures. Thus, achieving dense and stable 3D acquisition for fast-moving and deformable objects remains an open problem. We propose a technique to achieve dense shape reconstruction that requires only a single-frame image of a grid pattern based on coplanarity constraints. With our technique, positional information is not encoded in local regions of a projected pattern, but is distributed over the entire grid pattern, which results in robust image processing and 3D reconstruction. The technique also has the advantage of low computational cost due to its efficient formulation.

1. Introduction

Dense and accurate 3D scanning of dynamic objects, such as human facial expressions, or robust 3D scanning of dynamic scenes with moving sensors, such as autonomous robots' vision systems, are essential for many applications. Since passive stereo techniques have difficulty in reconstructing textureless surfaces densely and accurately, active 3D measurement techniques, especially those using

high-speed structured light systems, have been extensively studied in recent years.

Many structured light systems temporally encode positional information about a projector's pixel into multiple patterns. Recently, structured light systems that can capture a dynamic scene have been proposed^(6),19). These systems work by reducing the required number of patterns and increasing pattern speed. However, they assume that there is little motion in a scene while a sufficient number of patterns are projected and that target objects are rigid bodies. In addition, the design of high-speed synchronization systems is also an issue.

On the other hand, 'one-shot' structured light techniques using only a single image in which positional information of the projectors' pixels is embedded into spatial patterns of the projected image have also been studied^(8),13),15),17). Although these techniques can resolve the issues of rapid motion and synchronization, they typically use complex patterns of intensities or colors to encode positional information into local areas. Because of the complex patterns, assumptions of a smooth surface or reflectance are often required, and the image processing tends to be difficult: if the assumptions do not hold, the decoding process may lead to unstable reconstruction.

In this paper, we present a single-image scanning technique resolving the aforementioned problems. The proposed technique uses a simple grid pattern formed by a number of straight lines distinguishable only as vertical or horizontal lines so that image processing is simple and stable. In addition, there is no need to encode particular information about the local grid pattern itself, so the pattern can be dense as long as the lines are extractable. In the framework of coded structured light, a shape cannot be reconstructed from such a pattern. Thus, a new technique that reconstructs the grid pattern using coplanarity constraints is proposed. The technique simultaneously decodes positional information of all the grid points that are connected by using constraints on coplanarity obtained from the detected positions of the grid points and the connectivity between them. Our method has several features which are important for real-world applications. Since the technique requires only local information of connectivity between adjacent grid points, the shape can be restored even when there are abrupt changes in depth due to an occlusion or in color due to texture. It also has an efficient formulation which decreases computational cost.

^{†1} Faculty of Information Sciences, Hiroshima City University

^{†2} Faculty of engineering, Saitama University

^{†3} Institute of Scientific and Industrial Research, Osaka University

This paper is organized as follows. Related work is discussed in Section 2, and the basic theory of shape reconstruction from a single grid pattern is described in Section 3. In Section 4, details for pattern configuration and techniques for the detection of projected patterns from captured images are described. The experimental results and evaluations are presented in Section 5. Finally, a discussion of the results concludes the paper.

2. Related work

Shape reconstruction techniques that encode positional information of a projector into temporal or spatial changes in a projected pattern have been widely investigated^(1),19). Techniques using only temporal changes are easy to implement, accurate, dense and robust, so they are commonly used thus far^(2),3),7). Since this technique needs multiple patterns for decoding, it is not suitable for high-speed capture.

Several methods for reducing the required number of patterns using both temporal and spatial changes have been presented^(6),19). Hall-Holt et al. proposed an enhanced method to eliminate the limitation of static scenes by tracking projected stripe boundaries and using the temporal and spatial coherence of the scene for 3D reconstruction⁽⁶⁾. However, the technique is still not suited to fast moving objects. Young et al. proposed an efficient method to reduce patterns by using multiple cameras⁽¹⁹⁾. Using multiple cameras is sometimes a problem because synchronization is needed. Weise et al. proposed a 3D scanning system composed of three cameras and a DLP video projector. The method uses a projector that had been specially modified to project multiple images with phase-shift temporal coding at high frequencies, and they demonstrated the use of phase-shift codes while correcting the distortion of the measured objects⁽¹⁸⁾.

Techniques using only spatial encoding of a pattern allow scanning with only a single-frame image^(8),13),15). These typically use complex patterns or colors for the decoding process and require assumptions of a smooth or continuous surface or assumptions of uniform or smooth reflectance, either locally or globally. If such assumptions do not hold true, which is usually the case in the real environments, the decoding process of the patterns is easily affected and this leads to ambiguities near depth or color discontinuities. Therefore, 3D reconstruction tends to be

unstable or fails in realistic situations.

Although it does not strictly involve a structured light system, methods for shape reconstruction of dynamic objects by spatio-temporal stereo matching have been proposed^(4),20),21). With these techniques, a projector is only used to provide a texture that changes over time. This allows a pair of stereo cameras to achieve high-quality depth reconstructions. However, they still require several patterns for identification and are not suitable for fast moving deformable objects. In addition, the techniques require either spatial or temporal synchronization.

Proesmans et al. used a single, black grid pattern for an active stereo system⁽¹⁴⁾. Using a white background, they captured textures simultaneously with the scanning. Their technique assumes orthographical projection of the pattern, so only a relative 3D shape within a continuous region can be obtained. Koninckx et al. have proposed a technique allowing dense shape reconstruction based on a single image using a simple set of stripes^(10),11). This was achieved by combining dense unidentified stripes and several identified stripes. Their method depends on relative numbering of the dense patterns, which assumes local smoothness of the surface and may be disturbed by shape discontinuities and line detection failures.

Furukawa et al. used multiple images projected by an uncalibrated line laser to construct simultaneous linear equations which are solved to reconstruct a scene or shape^(5),9). Such equations are derived from coplanar constraints of intersection points between temporally accumulated projected lines.

In our method, a grid pattern is projected onto the target scene and simultaneous equations are constructed from the intersections of the lines on the grid using the formulation of Furukawa et al. The projected lines are identified by solving the equations. In the proposed method, no initial identification of the grid lines is required and the lines can be dense as long as they can be extracted to acquire high-resolution shapes. In addition, since discrimination of just two colors on a captured image is required, the image processing becomes robust and stable. Unlike previous ‘one-shot’ structured light techniques, our technique does not assume smoothness or uniform reflectance of surfaces and is robust against local discontinuities of grid patterns. Thus, our method is more practical for real applications.

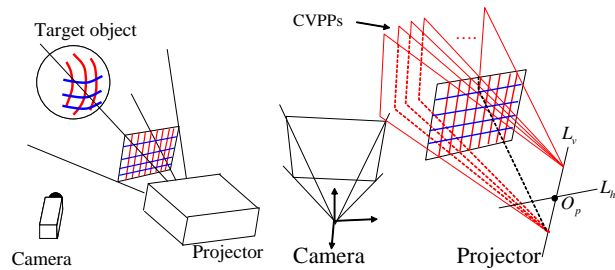


Fig. 1 Scanning system:(left) the system configuration, and (right) the definition of the coordinates.

3. Theory of shape reconstruction from a grid pattern

3.1 System configuration

The 3D measurement system consists of a camera and a projector as shown in Fig. 1(left). 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 projector pattern does not change, so no synchronization is needed. A grid pattern of vertical and horizontal lines is projected onto the scene and captured by the camera. The vertical and horizontal lines are assumed to be distinguishable. One way to achieve this is by using different colors for the vertical and horizontal lines and classifying them by color.

3.2 Problem definition

A straight line on the projected grid defines a plane in 3D space. Planes defined by a vertical line and a horizontal line are respectively termed a “Vertical Pattern Plane” (VPP) and a “Horizontal Pattern Plane” (HPP).

The projector is assumed to have been calibrated. Thus, all parameters for the VPPs and HPPs in 3D space are known. A VPP and a HPP with known parameters are termed a “Calibrated VPP” (CVPP) and a “Calibrated HPP” (CHPP), respectively. As is apparent, all the CVPPs share a common line L_v , as shown in Fig. 1(right). Note that this constraint is valid for any VPPs even if those are not identified, and thus, it is used for the formulation in Sections 3.4 and 3.5.1. Similarly, all the CHPPs share a common line L_h . The intersection of these two lines, L_v and L_h , corresponds to the optical center of the projector O_p .

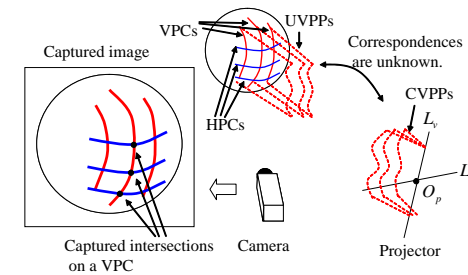


Fig. 2 CVPPs and UVPPs.

The point O_p and the direction vectors for L_v and L_h are known by calibration.

A vertical line projected onto the scene produces an observable 3D curve on the surface of the scene, and the curve is on the VPP defined by the line. In this paper, the 3D curve is termed a “Vertical Pattern Curve” (VPC). A “Horizontal Pattern Curve” (HPC) is defined in the same way. Intersections between the VPCs and HPCs are extracted from images captured by the camera. These points are termed “captured intersections.” The intersections are connected by the VPCs and HPCs and the connectivity is extracted by image processing (Fig. 2). Since the correspondence from each VPC detected in the image to a particular CVPP is unknown, a VPP that contains a VPC is termed an “Unknown VPP” (UVPP). An “Unknown HPP” (UHPP) is similarly defined. Note that CVPPs and CHPPs are not related to observation, but are determined only from calibration data, whereas UVPPs and UHPPs are associated by a one-to-one mapping to VPCs and HPCs that are observed in the image. Multiple UVPPs may be generated from a single projected line of the grid: for example, when the projected line is observed as multiple VPCs because of disconnections caused by noise, occlusion, or shape discontinuities.

The goal of the problem is to determine correspondences between the UVPPs (UHPPs) and CVPPs (CHPPs) (otherwise described as identifying UVPPs and UHPPs). As a result, 3D positions of all the captured intersections become known. Multiple UVPPs may correspond to a single CVPP if they are generated from a single projected line as mentioned above. The same assumption holds for UHPPs as well.

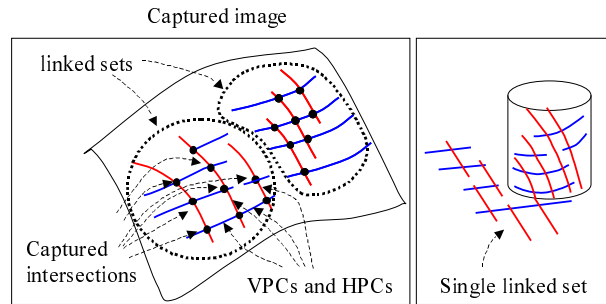


Fig. 3 Linked sets of captured intersections:(left) an example of a linked set, and (right) an example of a linked set that include discontinuity of the pattern.

A set of intersection points that are connected by VPCs or HPCs are termed “a linked set”. A linked set can include discontinuities between adjacent grid points as long as the two points are connected by multiple VPCs and HPCs as shown in Fig. 3. All the captured intersections are divided into multiple linked sets (e.g., two linked sets can be seen in Fig. 3(left)). In the proposed method, the target set of intersections is assumed to be contained in the same linked set. When there are multiple linked sets, each can be reconstructed using the same technique, so this assumption does not restrict the generality of the solution.

3.3 Outline of the solution

Fig. 4 shows an outline of the reconstruction processes of the proposed method. First, 2D sets of VPCs and HPCs are extracted from the image. As described in the previous section, UVPPs (or UHPPs) are associated to VPCs (or HPCs) by a one-to-one mapping. Thus, a UVPP is prepared for each VPC. UHPPs are prepared in the same way. Captured intersections are obtained by extracting intersection points between VPCs and HPCs, and simultaneous linear equations of UVPPs and UHPPs are obtained from these intersections. As will be shown in Section 3.4, the solution of UVPPs and UHPPs that are obtained by solving the linear equations has a one degree of freedom (1-DOF) indeterminacy. This indeterminacy can be resolved by matching the solution of UVPPs and UHPPs with the set of CVPPs and CHPPs obtained from the calibration parameters. By applying triangulation to the unique solution of UVPPs and UHPPs, 3D

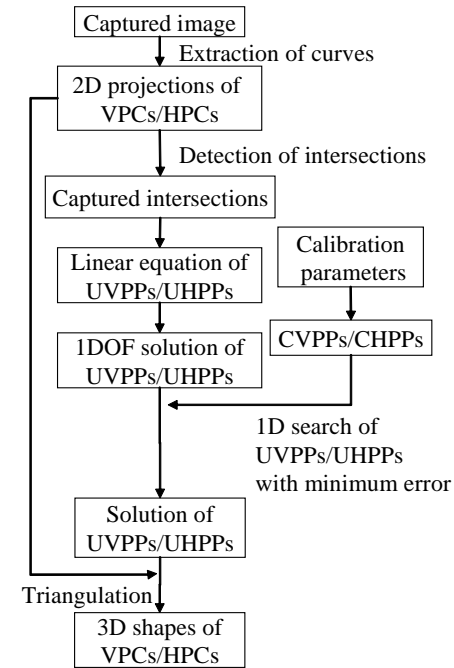


Fig. 4 Outline of the solution.

reconstruction of VPCs and HPCs is achieved.

The method for generating the simultaneous linear equations of UVPPs and UHPPs will be described in the following sections. As preparation for it, we first define symbols for UVPPs, UHPPs, CVPPs, CHPPs and the relationships between them. Let the M CVPPs obtained by calibration be represented as V_1, V_2, \dots, V_M , and let the N CHPPs be represented as H_1, H_2, \dots, H_N . Also, let the m UVPPs and n UHPPs obtained from the captured image be represented as v_1, v_2, \dots, v_m and h_1, h_2, \dots, h_n , respectively. As noted in Section 3.3, a single vertical line projected may be detected as multiple VPCs. Thus, m (or n) may be greater than M (or N).

The proposed method derives linear equations based on conditions of coplanarity with regard to UVPPs and UHPPs; that is, a captured intersection pro-

vides a linear constraint equation with regard to the UVPP or UHPP that contains it. In addition, all UVPPs should include L_v . Because of this, two linear constraint equations are obtained with regard to each of the UVPPs and UHPPs. These equations form a system of linear equations and can be solved by existing methods, such as singular value decomposition (SVD). In the case that the captured intersections are included in a linked set, this equation typically has a single trivial solution and retains one degree of freedom; as it stands, it will not yield a unique solution.

Determining the 1-DOF parameter for the solution of UVPPs and UHPPs is achieved by a 1D optimization: we find the parameter that minimizes a target function representing the errors between UVPPs/UHPPs and CVPPs/CHPPs. Since minimization of a target function with a continuous parameter is not easy, we discretize the parameter, and search for the optimum value from a finite set of values. To discretize the parameter, we sample the parameters where a certain UVPP (for example, v_1) coincides with any of the CVPP (V_i for $1 \leq i \leq M$). This process will be described in Section 3.6.

3.4 Solving coplanarity constraints

In the rest of this paper, coordinates of vectors or points are given in the camera coordinate system, whose origin is at the optical center of the camera. The x -axis is directed toward the right of the camera, the y -axis is directed down, and z -axis is directed toward the front.

From the intersections of the VPCs and HPCs obtained from the captured image, linear equations can be derived. Suppose that the intersection between v_k and h_l is captured and its 2D position on the image in the coordinates of the normalized camera is $\mathbf{u}_{k,l} = [s_{k,l}, t_{k,l}]^\top$. We assume that no planes include the optical center of the camera. Then, planes v_k and h_l are represented by

$$\mathbf{v}_k^\top \mathbf{x} = -1, \quad \mathbf{h}_l^\top \mathbf{x} = -1, \quad (1)$$

respectively, where 3D vectors \mathbf{v}_k and \mathbf{h}_k represent plane parameters and \mathbf{x} represents an arbitrary point on the planes ^{*1}. Let the 3D location of intersection

^{*1} Generally, a plane with normal vector \mathbf{n} and with signed distance d from the origin can be represented by $\mathbf{n}^\top \mathbf{x} = d$. From this representation, Equation (1) can be obtained by, for example, defining $\mathbf{v}_k \equiv -\mathbf{n}/d$. Note that $d \neq 0$ because the plane does not include the origin.

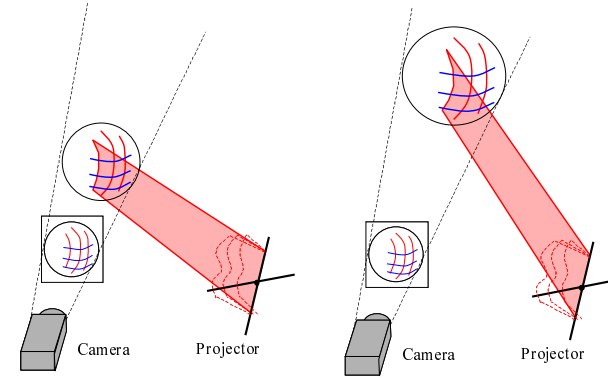


Fig. 5 1-DOF indeterminacy similar to scaling ambiguity.

$\mathbf{u}_{k,l}$ be represented as $\mathbf{x}_{k,l}$. Then, $\mathbf{x}_{k,l}$ can be represented using $\mathbf{u}_{k,l}$ as

$$\mathbf{x}_{k,l} = \gamma_{k,l} [\mathbf{u}_{k,l}^\top \quad 1]^\top, \quad (2)$$

where $\gamma_{k,l}$ is the depth value of the intersection. By eliminating $\mathbf{x}_{k,l}$ and $\gamma_{k,l}$ from Equations (1) and (2), we obtain

$$[\mathbf{u}_{k,l}^\top \quad 1] (\mathbf{v}_k - \mathbf{h}_l) = 0. \quad (3)$$

In addition, v_k contains line L_v which passes through the optical center O_p of the projector, as shown in Fig. 2. Let \mathbf{o}_p be the 3D coordinates of O_p and \mathbf{l}_v be the direction vector of L_v . Then, since v_k includes point O_p and line L_v , all the UVPPs v_k pass through \mathbf{o}_p , and their normal vectors are orthogonal to L_v . Thus,

$$\mathbf{o}_p^\top \mathbf{v}_k = -1, \quad \mathbf{l}_v^\top \mathbf{v}_k = 0. \quad (4)$$

Similarly, all the UVPPs h_l pass through \mathbf{o}_p , and their normal vectors are orthogonal to L_h . Thus, with the direction vector for L_h being \mathbf{l}_h , we have

$$\mathbf{o}_p^\top \mathbf{h}_l = -1, \quad \mathbf{l}_h^\top \mathbf{h}_l = 0. \quad (5)$$

By putting Equations (3), (4) and (5) together, the system of linear equations

$$\mathbf{P}\mathbf{q} = \mathbf{b} \quad (6)$$

where $\mathbf{q} = [\mathbf{v}_1^\top, \dots, \mathbf{v}_m^\top, \mathbf{h}_1^\top, \dots, \mathbf{h}_n^\top]^\top$ is obtained.

If all the UVPPs and UHPPs coincide with the plane that contains both L_v and L_h , the aforementioned linear equation holds. This trivial solution is denoted by \mathbf{w} . If we determine the depth of one captured intersection, the depths of all

the intersections connected to it can be found because connected intersections are on a single plane. Thus, the general solution \mathbf{q} of Equation (6) has a 1-DOF indeterminacy (see Fig. 5). Since \mathbf{w} is a particular solution and \mathbf{q} has 1-DOF indeterminacy,

$$\mathbf{q} = \mathbf{w} + p\mathbf{u}, \tag{7}$$

where p is an arbitrary scalar value and \mathbf{u} is a non-zero vector that satisfies $\mathbf{P}\mathbf{u} = \mathbf{0}$. There should be such a \mathbf{u} if there is non-trivial solution $\mathbf{q} \neq \mathbf{w}$. Equation (7) can be confirmed to be a general solution because $\mathbf{P}(\mathbf{w} + p\mathbf{u}) = \mathbf{P}\mathbf{w} + p\mathbf{P}\mathbf{u} = 0 + p0$ from the definitions of \mathbf{w} and \mathbf{u} . Here, \mathbf{w} and \mathbf{u} can be calculated by SVD of \mathbf{P} as follows. Using SVD, we can obtain the decomposition $\mathbf{P} = \mathbf{U} \text{diag}(w_1, w_2, \dots, w_{3(m+n)}) \mathbf{V}^\top$, where \mathbf{U} and \mathbf{V} are matrices with column vectors that are orthogonal to each other, and $w_1, w_2, \dots, w_{3(m+n)}$ are the singular values sorted in decreasing order. From the SVD result, \mathbf{u} in Equation (7) can be obtained as the right-most column of \mathbf{V} , and a particular solution \mathbf{w} is given by $\mathbf{w} = \mathbf{V} \text{diag}(1/w_1, 1/w_2, \dots, 1/w_{3(m+n)-1}, 0) \mathbf{U}^\top \mathbf{b}$.

3.5 Efficient calculation by reducing variables

3.5.1 1-parameter representation of planes

The method for obtaining general solutions of planes v_1, \dots, v_m , and h_1, \dots, h_n described in the previous section requires solving a linear equation (6) which has $3(m+n)$ variables. If dense CVPPs (or either dense CHPPs) are used to increase the density of the measured points, m or n may be a large number. Thus, the computational complexity of reconstruction may become computationally complex. In this section, techniques for reducing the number of variables in Equation (6) are described. By using these techniques, reconstruction can be achieved with much less computational complexity.

First we describe a method for reducing $3(m+n)$ variables to $m+n$ variables. Let \mathbf{o}_p be the optical center of the projector. Let \mathbf{p} be the parameter vector of the plane that passes through \mathbf{o}_p and includes both L_v and L_h . Thus, planes v_k are elements of a pencil of planes that share L_v , where L_v is a line that passes through \mathbf{o}_p with direction vector \mathbf{l}_v . Then, the parameter vector \mathbf{v}_k of the plane v_k can be written as $\mathbf{v}_k = \eta\bar{\mathbf{v}} + \mathbf{s}$, where η is a scalar value, $\bar{\mathbf{v}} \equiv \mathbf{l}_v \times \mathbf{o}_p$, and

\mathbf{s} is a parameter vector of an arbitrary element of the pencil of planes ^{*1}. This parameter vector \mathbf{v}_k parameterizes the pencil of planes with a single variable η . Similarly, \mathbf{h}_l can be written as $\mathbf{h}_l = \rho\bar{\mathbf{h}} + \mathbf{t}$, where ρ is a scalar value, where $\bar{\mathbf{h}}$ is defined as $\bar{\mathbf{h}} \equiv \mathbf{l}_h \times \mathbf{o}_p$, and \mathbf{t} is a parameter vector of an arbitrary element of the pencil of planes that share L_h . Note that the pencil of planes that share L_v and the pencil of planes that share L_h have a unique common plane, \mathbf{p} , and the vector \mathbf{p} can be obtained from calibration. Thus, \mathbf{v}_k and \mathbf{h}_l can be given by the 1-parameter representations as

$$\mathbf{v}_k = \eta_k\bar{\mathbf{v}} + \mathbf{p}, \mathbf{h}_l = \rho_l\bar{\mathbf{h}} + \mathbf{p}. \tag{8}$$

Let $\tilde{\mathbf{u}}_{k,l}$ be defined as $\tilde{\mathbf{u}}_{k,l} \equiv [\mathbf{u}_{k,l}^\top \ 1]^\top$. Then, from Equation (3),

$$\begin{aligned} \tilde{\mathbf{u}}_{k,l}^\top \{(\eta_k\bar{\mathbf{v}} + \mathbf{p}) - (\rho_l\bar{\mathbf{h}} + \mathbf{p})\} &= 0 \\ \Leftrightarrow (\tilde{\mathbf{u}}_{k,l}^\top \bar{\mathbf{v}})\eta_k - (\tilde{\mathbf{u}}_{k,l}^\top \bar{\mathbf{h}})\rho_l &= 0. \end{aligned} \tag{9}$$

We define constants $F_{k,l} \equiv \tilde{\mathbf{u}}_{k,l}^\top \bar{\mathbf{v}}$ and $G_{k,l} \equiv \tilde{\mathbf{u}}_{k,l}^\top \bar{\mathbf{h}}$. Then, Equation (9) can be written as

$$F_{k,l}\eta_k = G_{k,l}\rho_l. \tag{10}$$

By accumulating Equation (10) for all the captured intersections, q simultaneous equations with $(m+n)$ variables (variables are $\eta_k, \rho_l, 1 \leq k \leq m, 1 \leq l \leq n$) are obtained, where q is the number of captured intersections.

Let t be the index of a captured intersection, and let $\iota(t)$ and $\kappa(t)$ be the indices of the vertical and horizontal planes that go through captured intersection t (e.g., if v_3 and h_4 cross at the 5-th captured intersection, $\iota(5) = 3$ and $\kappa(5) = 4$). Let \mathbf{T} and \mathbf{R} be a $q \times m$ matrix and a $q \times n$ matrix where their (r, c) elements $\mathbf{T}_{r,c}$ and $\mathbf{R}_{r,c}$ are

$$\mathbf{T}_{r,c} = \begin{cases} F_{\iota(r),\kappa(r)} & \text{if } c = \iota(r) \\ 0 & \text{otherwise} \end{cases}, \mathbf{R}_{r,c} = \begin{cases} G_{\iota(r),\kappa(r)} & \text{if } c = \kappa(r) \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

Then, by defining $\vec{\eta} \equiv (\eta_1, \dots, \eta_m)^\top$ and $\vec{\rho} \equiv (\rho_1, \dots, \rho_n)^\top$, the simultaneous equations (10) can be represented as

$$\mathbf{T}\vec{\eta} = \mathbf{R}\vec{\rho}. \tag{12}$$

^{*1} Note that Equation (4) holds for such a plane, because $\mathbf{o}_p^\top(\eta\bar{\mathbf{v}} + \mathbf{s}) = \eta\{\mathbf{o}_p^\top(\mathbf{l}_v \times \mathbf{o}_p)\} + \mathbf{o}_p^\top \mathbf{s} = \eta \cdot 0 + (-1) = -1$ ($\mathbf{o}_p^\top \mathbf{s} = -1$, since plane \mathbf{s} includes point \mathbf{o}_p) and $\mathbf{l}_v^\top(\eta\bar{\mathbf{v}} + \mathbf{s}) = \eta\{\mathbf{l}_v^\top(\mathbf{l}_v \times \mathbf{o}_p)\} + \mathbf{l}_v^\top \mathbf{s} = \eta \cdot 0 + 0 = 0$ ($\mathbf{l}_v^\top \mathbf{s} = 0$, since plane \mathbf{s} includes L_v).

To solve Equation (12) by the least squares method,

$$\|\mathbf{T}\vec{\eta} - \mathbf{R}\vec{\rho}\|^2 = \left\| [\mathbf{T} | -\mathbf{R}] \begin{bmatrix} \vec{\eta} \\ \vec{\rho} \end{bmatrix} \right\|^2 \quad (13)$$

should be minimized. This can be achieved by calculating the eigenvector associated with the minimum eigenvalue of the symmetric $(m+n) \times (m+n)$ matrix $[\mathbf{T} | -\mathbf{R}]^\top [\mathbf{T} | -\mathbf{R}]$. There are efficient numerical algorithms for this problem. In this mathematical representation, the 1-DOF indeterminacy of the equations corresponds to the scaling ambiguity of the eigenvector.

3.5.2 Reducing variables from dependency of variables

It is possible to reduce the number of variables in Equation (13) further. A least squares solution is given by

$$\min_{\vec{\eta}, \vec{\rho}} \|\mathbf{T}\vec{\eta} - \mathbf{R}\vec{\rho}\|^2 = \min_{\vec{\rho}} (\min_{\vec{\eta}} \|\mathbf{T}\vec{\eta} - \mathbf{R}\vec{\rho}\|^2). \quad (14)$$

The solution of $\min_{\vec{\eta}} \|\mathbf{T}\vec{\eta} - \mathbf{R}\vec{\rho}\|^2$ with a fixed $\vec{\rho}$ is achieved when $\vec{\eta} = \mathbf{T}^\dagger \mathbf{R}\vec{\rho}$, where $\mathbf{T}^\dagger \equiv (\mathbf{T}^\top \mathbf{T})^{-1} \mathbf{T}^\top$ is a pseudo inverse of \mathbf{T} .

The (r, c) element of $\mathbf{T}^\top \mathbf{T}$, which is described as $(\mathbf{T}^\top \mathbf{T})_{r,c}$, is

$$(\mathbf{T}^\top \mathbf{T})_{r,c} = \sum_w \mathbf{T}_{r,w}^\top \mathbf{T}_{w,c} = \sum_w \mathbf{T}_{w,r} \mathbf{T}_{w,c}. \quad (15)$$

This means that $(\mathbf{T}^\top \mathbf{T})_{r,c}$ is the inner product of the r -th column and the c -th column of the matrix \mathbf{T} . Since each row of \mathbf{T} has only one non-zero element from the definition of \mathbf{T} , different columns of \mathbf{T} are always orthogonal. Thus, $\mathbf{T}^\top \mathbf{T}$ is a diagonal matrix where

$$(\mathbf{T}^\top \mathbf{T})_{r,c} = \begin{cases} \sum_w (\mathbf{T}_{w,r})^2 & \text{if } r = c \\ 0 & \text{otherwise} \end{cases} \quad (16)$$

Thus, $(\mathbf{T}^\top \mathbf{T})^{-1}$ can be directly calculated and \mathbf{T}^\dagger is obtained by simple multiplication.

Substituting $\vec{\eta} = \mathbf{T}^\dagger \mathbf{R}\vec{\rho}$ into Equation (14), we obtain

$$\min_{\vec{\eta}, \vec{\rho}} \|\mathbf{T}\vec{\eta} - \mathbf{R}\vec{\rho}\|^2 = \min_{\vec{\rho}} (\|\mathbf{T}\mathbf{T}^\dagger \mathbf{R}\vec{\rho} - \mathbf{R}\vec{\rho}\|^2) \quad (17)$$

$$= \min_{\vec{\rho}} (\|(\mathbf{T}\mathbf{T}^\dagger \mathbf{R} - \mathbf{R})\vec{\rho}\|^2). \quad (18)$$

This means that the optimal value of $\vec{\rho}$, which we denote by $\hat{\rho}$, can be calculated as a unit eigenvector associated with the minimum eigenvalue of the matrix $(\mathbf{T}\mathbf{T}^\dagger \mathbf{R} - \mathbf{R})^\top (\mathbf{T}\mathbf{T}^\dagger \mathbf{R} - \mathbf{R})$. Then the optimal value $\hat{\eta}$ is obtained by $\hat{\eta} = \mathbf{T}^\dagger \mathbf{R}\hat{\rho}$. Since $(\mathbf{T}\mathbf{T}^\dagger \mathbf{R} - \mathbf{R})^\top (\mathbf{T}\mathbf{T}^\dagger \mathbf{R} - \mathbf{R})$ is a $n \times n$ symmetric matrix, the computational

complexity can be further reduced.

Let $\hat{\eta}_i$ be the i -th component of the optimal solution vector $\hat{\eta}$, and $\hat{\rho}_j$ be the j -th component of $\hat{\rho}$. Then, the general solution of UVPPs and UHPPs is

$$\mathbf{v}_k(\lambda) \equiv \lambda \hat{\eta}_k \bar{\mathbf{v}} + \mathbf{p}, \quad \mathbf{h}_l(\lambda) \equiv \lambda \hat{\rho}_l \bar{\mathbf{h}} + \mathbf{p}, \quad (19)$$

where λ is the parameter for the 1-DOF indeterminacy.

3.6 Resolving the 1-DOF ambiguity

The general solution of UVPPs and UHPPs in Equation (19) includes an unknown scalar λ . The scalar λ can be determined by matching the UHPPs (or UVPPs) with the parameter λ to the CHPPs (or CVPPs) which are known from the calibration data. Comparisons are made between UVPPs $\mathbf{v}_k(\lambda)$ and CVPPs \mathbf{V}_i and also between UHPPs $\mathbf{h}_l(\lambda)$ and CHPPs \mathbf{H}_j , where \mathbf{V}_i and \mathbf{H}_j are the parameter vectors of CVPPs and CHPPs. In this paper, the comparison is based on the squared angles between the planes. More specifically, the error function is defined as

$$E(\lambda) \equiv \sum_{k=1}^m \min_{i=1, \dots, M} \{D(\mathbf{v}_k(\lambda), \mathbf{V}_i)\}^2 + \sum_{l=1}^n \min_{j=1, \dots, N} \{D(\mathbf{h}_l(\lambda), \mathbf{H}_j)\}^2, \quad (20)$$

where

$$D(\mathbf{v}_k, \mathbf{V}_i) \equiv \arccos\left(\frac{\mathbf{v}_k \cdot \mathbf{V}_i}{\|\mathbf{v}_k\| \|\mathbf{V}_i\|}\right) \quad (21)$$

is the angle between two planes, \mathbf{v}_k and \mathbf{V}_i . Then, the optimal value of λ is given by

$$\lambda^* \equiv \arg \min E(\lambda). \quad (22)$$

Then, the set of planes $\mathbf{v}_k(\lambda^*)$, ($k = 1, 2, \dots, m$) and $\mathbf{h}_l(\lambda^*)$, ($l = 1, 2, \dots, n$) is the solution.

The error function (20) represents the differences between the entire set of CVPPs (or CHPPs) and the set of UVPPs (or UHPPs) reconstructed from the captured intersections. Thus, the proposed method uses information of the locations of all the CVPPs and CHPPs simultaneously to resolve the ambiguity in the coplanarity constraints.

To reduce the search domain of Equation (22), we assume that the UVPPs must match to the CVPPs. Based on this assumption, a predefined UVPP that is selected in advance (a simple selection method is using the first UVPP $\mathbf{v}_1(\lambda^*)$) should coincide with one of the CVPPs for the optimum solution with

the parameter λ^* . More specifically, if the first UVPP $\mathbf{v}_1(\lambda^*)$ coincides with the i -th CVPP \mathbf{V}_i , the following constraint should be satisfied.

$$\mathbf{V}_i = \mathbf{v}_1 = \lambda(1, i) \tilde{\eta}_1 \bar{\mathbf{v}} + \mathbf{p}, \quad (23)$$

where $\lambda(k, i)$ is defined as the value of λ when UVPP $\mathbf{v}_1(\lambda^*)$ coincides with CVPP \mathbf{V}_i . Thus, $\lambda(1, i)$ can be calculated from $\lambda(1, i) = \|\mathbf{V}_i - \mathbf{p}\| / \|\tilde{\eta}_1 \bar{\mathbf{v}}\|$.

If the assumption described above is satisfied, λ^* should be an element of the set $\{\lambda(1, i) | 1 \leq i \leq M\}$. So, λ^* can be found by searching for the $\lambda(1, i)$ that gives the minimum value of Equation (22) with respect to $1 \leq i \leq M$. Thus, $\lambda^* = \arg \min_{1 \leq i \leq M} E(\lambda(1, i))$.

Theoretically, if there are no observation errors, the proposed method can reconstruct a linked set that consists of only a single intersection generated by a UVPP and a UHPP (the case of $m = n = 1$). This extreme case is equivalent to identifying the intersection using epipolar constraints. In actual cases, however, reconstruction with only a single intersection may become unstable because of noise and dense grid patterns. By using multiple UVPPs and UHPPs, tolerance to noise improves because more information is used in the error function (20).

4. Configuration and detection of the grid pattern

4.1 Configuration of the grid pattern

As described in Section 3.6, the stability of the search for λ^* is affected by the locations of CVPPs and CHPPs. If the error function (20) has a unique minimum value at the true solution, the correct solution is obtained by the search for λ^* . However, if the function has multiple minimum values, the search may fail. Fortunately, the cases of multiple minimum values (*e.g.*, the projector and the camera are parallel, the pattern is a square and uniform grid, and the optical center of the projector is on the line of $x = y, z = 0$ in the camera coordinates) seldom happen in real situations and the search normally succeeds.

Another problem is that error values near the minimum value are not usually significantly different from the minimum value and it is difficult to search for the true solution due to the presence of noise. A simple way to prevent this is by placing the CVPPs and CHPPs at irregular intervals on the projector's image plane. By doing this, the irregularity makes the minimum value of the energy function more distinguishable from other minimal values.

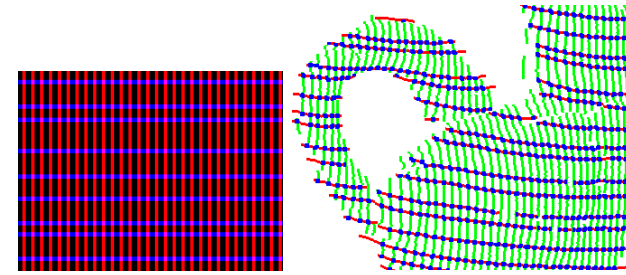


Fig. 6 Example of (left) a projected pattern and (right) detected HPCs and VPCs. In the right figure, red curves are HPCs, green curves are VPCs, and intersection points are blue dots.

To achieve a more robust search, a sparse pattern is preferable, whereas dense patterns are required for a detailed scan of an object. Therefore, it is difficult to satisfy both requirements at the same time. One effective solution is to combine patterns of dense vertical lines with uniform intervals and horizontal lines with irregular intervals. Random intervals are used in this paper (Fig. 6(left)).

In terms of pattern, any kind of pattern that consists of vertical and horizontal lines can be a candidate: *e.g.*, stripe patterns or checkerboard patterns. In our implementation, we adopt stripe patterns for simplicity. Vertical and horizontal stripes are colored red and blue, respectively, so that they were easily distinguishable.

Sub-pixel accuracy for the line detection is needed to obtain results with sufficient quality. We used peak detection with parabolic fitting. The stability and simplicity of the peak detection algorithm is another reason for adopting stripe patterns for our method (a checkerboard pattern requires an edge detection algorithm and it is not as stable as peak detection).

Note that our method does not directly encode information into the patterns, but uses their connectivity, and so complicated encoding patterns, like a Debruijn sequence⁸⁾, are not required. Because of these features, this method is less affected by distortion due to surface texture and object shape, resulting in robust and stable 3D reconstruction with simple image processing.

4.2 Detection of grid pattern

To project a grid pattern, we usually use a video projector as a light source.

Since the intensity of the light is relatively weak compared to strong light sources like lasers or special projectors, simple thresholding techniques sometimes fail. Therefore, we implemented a curve extraction algorithm based on peak detection to increase robustness.

First, the captured image is scanned horizontally and vertically, detecting VPCs and HPCs, respectively, with sub-pixel accuracy using the peak detection algorithm. For example, red VPCs are detected by horizontal scanning of an image. The intensity profiles in the red-channel of a single row of the pixels are examined and a sub-sequence of the signal where the intensity monotonically increases first and then monotonically decreases is extracted. Let the intensity profile of the j -th row be $f_r(i, j)$, ($i = 0, 1, 2, \dots$) and let the pixel position of the maximum intensity be (i_p, j) . If there exist integers k, l where $f(i_p, j) - f(i_p - k, j) \geq C_f$, $1 \leq k \leq C_w$ and $f(i_p, j) - f(i_p + l, j) \geq C_f$, $1 \leq l \leq C_w$ (C_f is the threshold of the minimum gradient of the steeps of the peaks and C_w is the threshold about the maximum width of the peaks), then (i_p, j) is detected as a peak position at pixel-accuracy. The peak position at sub-pixel accuracy is calculated as the maximum position of the quadratic function that includes $f(i_p - 1, j)$, $f(i_p, j)$ and $f(i_p + 1, j)$. Detected peaks of the j -th row are compared with peaks of $j - 1$ -th row, and if the distance to the nearest peak is smaller than a threshold, they are connected to form a VPC. Isolated peaks are then removed. In our experiments, this simple technique worked sufficiently well to retrieve satisfactory results. However, many line detection algorithms have already been proposed and may be used to obtain better results. Fig. 6(right) shows example of detected lines and intersection points.

In processing real images, several types of errors in pattern detection may occur. One of these is disconnection of detected VPCs or HPCs. This type of error is not a serious problem because the solution of the linear equation (6) remains the same as long as the intersections configure the same linked set.

Other types of incorrect connection of multiple VPCs (or HPCs) or incorrect detection of intersections changes the solution of the linear equation. If both of the wrongly connected lines are included in a single linked set, a false constraint is added to the system of simultaneous linear equations. Since the system of linear equations is normally over-constrained, our method is robust against this

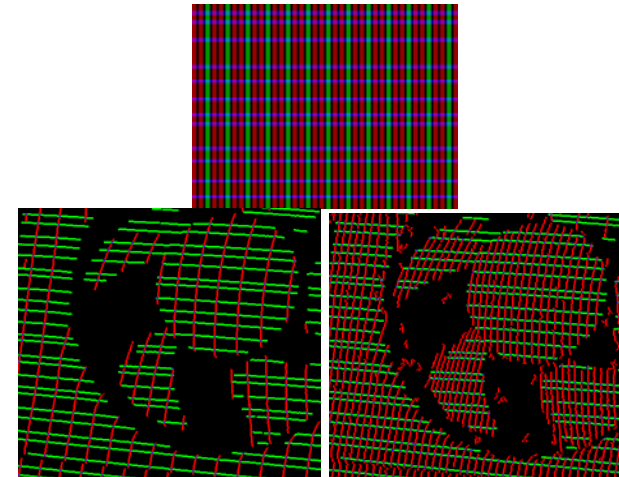


Fig. 7 A pattern for coarse to fine technique. (top) a projected pattern, (bottom left), detected coarse lines and (bottom right) detected fine lines. In the bottom figures, green curves are horizontal patterns, red curves are vertical, and intersection points are blue dots.

type of error.

When two different linked sets become wrongly connected, a major change in the result may occur. If one linked set is much larger than the other, the larger set tends to be reconstructed correctly whereas smaller set becomes distorted by a generalized projective bas-relief (GPBR) transformation¹²⁾, because the larger set dominates calculation of the error function. If both sets are about the same size, both reconstructed sets may be distorted.

The only requirements that should be achieved in the image processing step of the proposed method, is the extraction of the grid lines that are labeled as vertical or horizontal lines. This simplicity is a huge advantage for achieving robust image processing. In an experiment, which will be described in Section 5.4, the reconstruction of textured objects was demonstrated.

4.3 Dense reconstruction with coarse-to-fine technique

The frequency of the projected pattern is another problem. If we want to capture the shape more densely, a higher frequency pattern is required. However, if we set the frequency higher than the image resolution, the correct pattern can

not be captured because a number of lines merge into a single segment on the image and the shape can not be recovered correctly. Moreover, the ideal frequency of the projected pattern may be different depending on the location of the scene. For example, even if the frequency is ideal for the front face of the cylindrical shape, the frequency may become too high near the occluding boundary of the object, which may result in failure of reconstruction.

In this paper, we propose the following simple method to solve the problem as follows. We have already used two colors for identification of vertical and horizontal lines. Therefore, we use additional colors for different frequencies. In our implementation, we used only three colors for robust image processing. Therefore we project lines at two different frequencies (*e.g.*, dense and sparse patterns). Fig. 7 shows an example pattern, detected lines and intersection points. In the figure, we can see that both of the patterns were correctly detected. To merge the shapes of two patterns, we can simply add all the recovered points. More intelligent methods can be applied to improve results.

5. Experiments

5.1 Evaluation of accuracy and robustness using simulation data

Several experiments were conducted to test the proposed method.

The first experiment confirmed the validity of the proposed method using simulated data. Several grid patterns were provided to synthesize simulated images. The first grid consists of a complete set of lines aligned at uniform intervals. The second is the same as the first, except that the intervals of the horizontal lines are randomized to disturb the uniformity of the grid pattern. This increases the stability of the determination of correspondences, as described in Section 4.1. The synthesized camera images of the grid patterns are shown in Fig. 8. In the images, the intervals of the VPCs are about 5 pixels. The intersections of the grid patterns were extracted from the images and the correspondences from the UHPPs and UVPPs to the CVPPs and CHPPs were determined using the proposed method. For both patterns, the correct correspondences for all the UHPPs and UVPPs were selected and the reconstructed shapes exactly matched the ground truth. The shape obtained from the data of irregular (random) intervals is shown in Fig. 9 together with the ground truth.

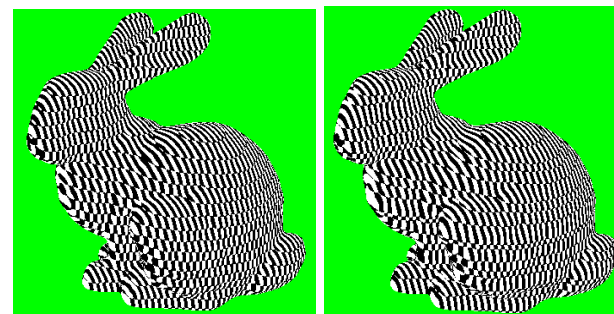


Fig. 8 Synthesized grid patterns:(left) uniform intervals and (right) irregular (randomized) intervals.

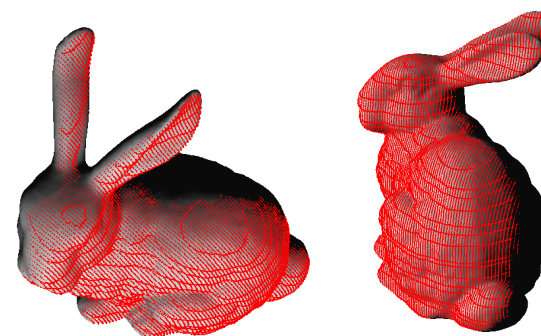


Fig. 9 Result of simulation data with irregular (randomized) intervals. The red points are reconstructed points and the shaded surface is the ground truth.

The processing time for reconstructing the input data created using the pattern with randomized intervals on a PC with a 3.8GHz Xeon CPU was about 1.6 sec. Almost all of the time was spent for the SVD calculation. Once \mathbf{w} and \mathbf{u} of Equation (7) had been obtained from the SVD result, the minimum value search took less than 0.001s.

Another experiment was conducted to evaluate the stability of the proposed method when the input data (the set of captured intersections) were disturbed by noise. Since the stability of the proposed method depends on the projected pattern, the two types of patterns shown in Fig. 8 were used and isotropic 2D

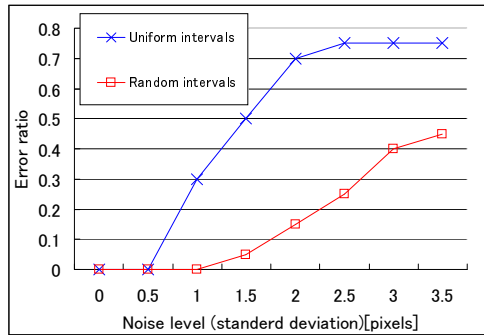


Fig. 10 Error ratios with different noise levels.

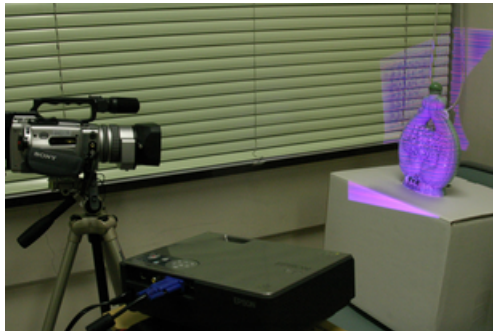


Fig. 11 Real 3D Scanning System

Gaussian noise was added to the captured intersections. The proposed method was applied to data with various noise levels (noise levels were defined by the standard deviation of the noise measured in pixels), and 20 tests were conducted at each noise level. For each test, correctness of the reconstruction was decided by whether the first UVPP of the solution ($v_1(\lambda^*)$) was associated to the correct CVPP or not. The ratio of correct reconstructions out of the 20 cases at each level is shown in Fig. 10. The results confirmed that stability of the algorithm was improved using the pattern with irregular (random) intervals.

5.2 Evaluation of the effect of reducing variables

In Section 3.5, a technique for reducing computational complexity by con-

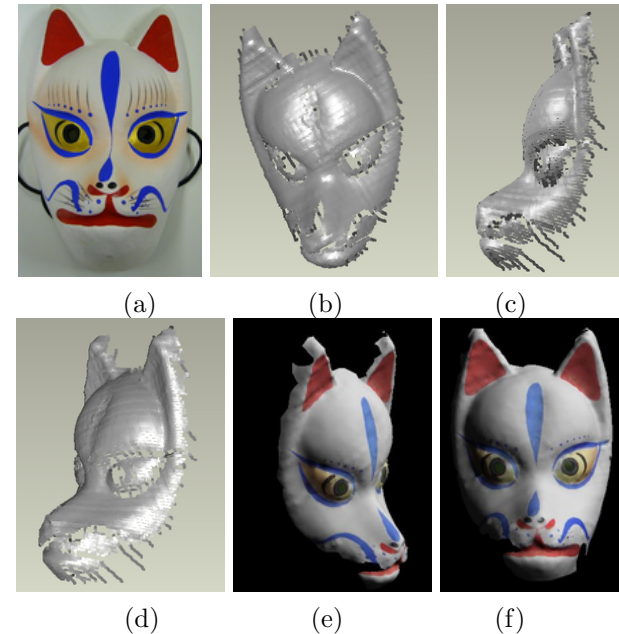


Fig. 12 Target object and reconstruction results: (a) the target objects, (b)-(d) reconstructed shape and (e),(f) textured shape after hole-filling.

structing linear equations with fewer variables is described. We conducted an experiment to demonstrate the effectiveness of this technique.

An actual 3D scanning system was built as shown in Fig. 11. Patterns were projected with a resolution of 1024×768 pixels and scenes were captured by a CCD camera (720×480 pixels). Fig. 12(a) shows the target object (a paper mask) used in this experiment. The target object was captured by the proposed system and its 3D shape was reconstructed by the following three methods: the method of Section 3.4 that uses $3(m + n)$ variables (method A); the method of Section 3.5.1 that uses $m + n$ variables (method B); and the method of Section 3.5.2 that uses n variables (method C). Then, the processing performances for the methods were compared.

Processing was performed by a PC with a 2.5GHz Xeon CPU and 2GB main

memory. Other conditions such as dimension of the image were the same as the previous experiment. The number of the detected vertical lines (*i.e.*, m) was 812, and the number of horizontal lines (n) was 142. The shape acquired by method C is shown in Figs. 12(b)-(f). We confirmed that the results of three methods were identical.

In Figs. 12(b)-(d), we can see noticeable traces on the results. This is mainly because of small errors in the normal vectors of points used for point shading. In our system, normal vectors are calculated using neighboring points. Since the proposed technique reconstructs shapes based on stripes, the density for the direction of stripes is very high, while the density for the orthogonal direction is not so high. Thus, if the kernel for normal calculation is isotropic, the accuracies of the normals for the former and the latter directions are different. This produces visual effects along the vertical direction.

The execution times were 104 seconds for method A (2862 variables), 3.5 seconds for method B (954 variables), and 0.0035 seconds for method C (142 variables). The number of variables of the linear equations was reduced by a factor of about 20 for method C compared to method A, whereas the execution time was reduced by a factor of about 30000. This is because the computational complexity of calculating eigenvectors scales much more quickly than linearly with respect to the dimension of the matrix. This experiment demonstrated that the method of variable reduction is very effective in reducing the computational cost of reconstruction.

5.3 Evaluation of the coarse-to-fine technique

Next, a scene of a box (size: 0.4 m \times 0.3 m \times 0.3 m) and a cylinder (height: 0.2 m, diameter: 0.2 m) was measured to test the coarse-to-fine method using the pattern proposed in Section 4.3. The scene was also measured by an active measurement method using coded structured light¹⁶⁾ to establish the ground truth. Fig. 13(a) shows the original captured scene while (b) and (c) show the captured images where objects are projected with the proposed pattern. Figs. 13(d) to (f) show the detected curves and (g) and (h) show the intersections of horizontal and vertical curves for the coarse and fine pattern, respectively. Fig. 13(k) shows the merged points of sparse (i) and dense reconstruction (j), and Figures (l) to (n) show both the reconstruction (red and green points) and the

ground truth (polygon mesh). Fig. 13(e) shows an example of line detection with the dense pattern, where the detection failed around the right side of the cylinder because the density exceeds the limit of detection. On the other hand, in (f), detection succeeded in the same region. Detection of intersections of the lines (g) and reconstruction (i) failed for the right side of the shape with the dense pattern, whereas detection of intersections (h) and reconstruction (j) succeeded using the sparse pattern. However, density of the reconstructed points of the sparse pattern is low on the left side, which can be seen from figure (j). By combining all these results, the entire cylinder shape could be reconstructed with sufficiently high density as shown in figure (k). Although there were small differences between the reconstruction and the ground truth, the correct shape was densely and accurately restored. The RMS error of the reconstruction from the ground truth was 0.52mm, where the minimum and maximum coordinates of the bounding box of the measured data were $(-200mm, -190mm, 660mm)$ and $(110mm, 160mm, 880mm)$.

Another experiment for demonstrating the effects of the coarse-to-fine approach was conducted. The results (Fig. 14) show that the reconstruction using the sparse pattern works better for the left side of the column in the scene, whereas the dense pattern works better for the front face. We see that, by combining results from both patterns, shapes that have faces with various directions can be reconstructed with high accuracy and density.

5.4 Evaluation of line detection

Line detection performances were evaluated for the cases that (1) the grid pattern was blurred and that (2) the target object has a complex texture. Fig. 15 shows the results for case (1). Fig. 15(b) shows an image that was defocused by about 30cm, where the fine patterns were not detected on the front side and only the coarse patterns were detected and reconstructed. Fig. 15 (e) shows an image defocused by about 50cm, where all the patterns on the front side could not be detected and reconstruction just failed. However, for the left side of the object, reconstruction was succeeded because the defocused pattern was sharpened by the tilted plane.

Fig. 16 shows the results for case (2), where an object with many colors and patterns was scanned. Although the object had a complex texture with many

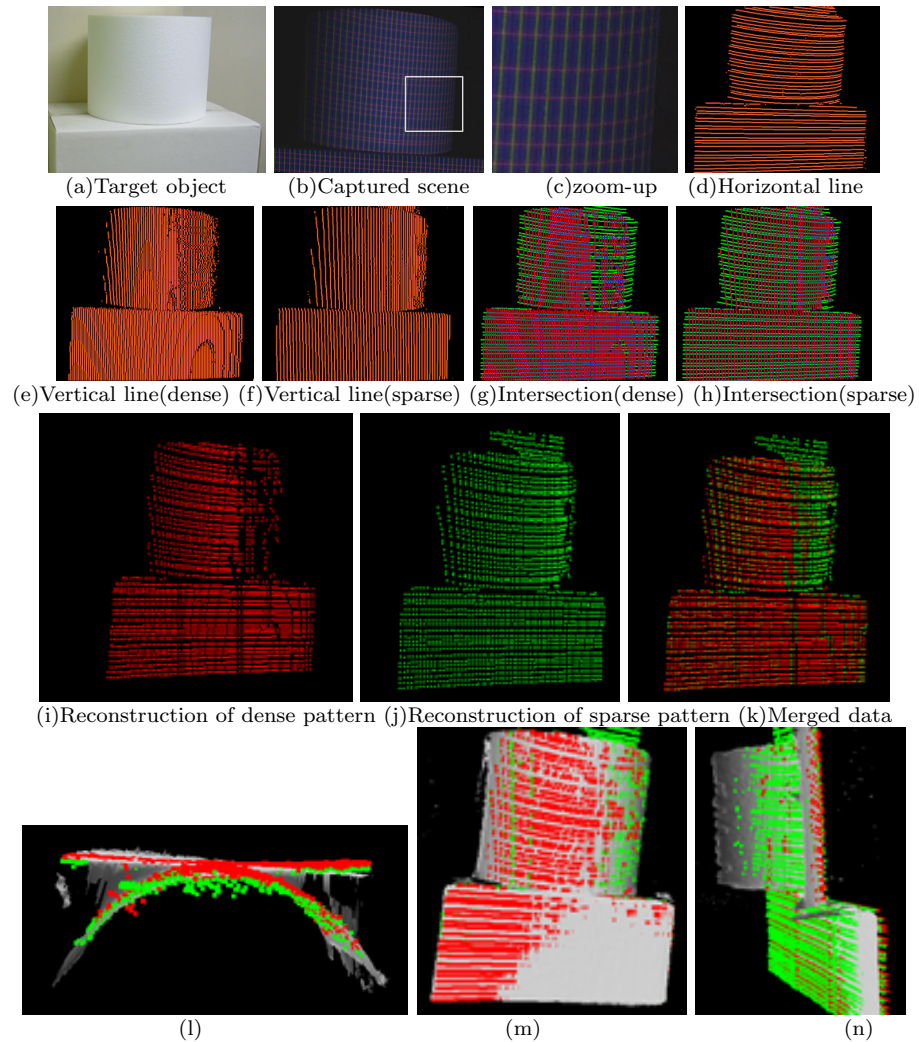


Fig. 13 Reconstruction and evaluation results: (a) captured scene, (b)(c) objects projected by the coarse-to-fine pattern, (d)-(f) detected curves, (g)(h) intersections of horizontal and vertical curves, (i)-(k) reconstruction results, and (l)-(n) the reconstructed model displayed with the ground truth data (red and green points: reconstructed model and shaded model: ground truth).

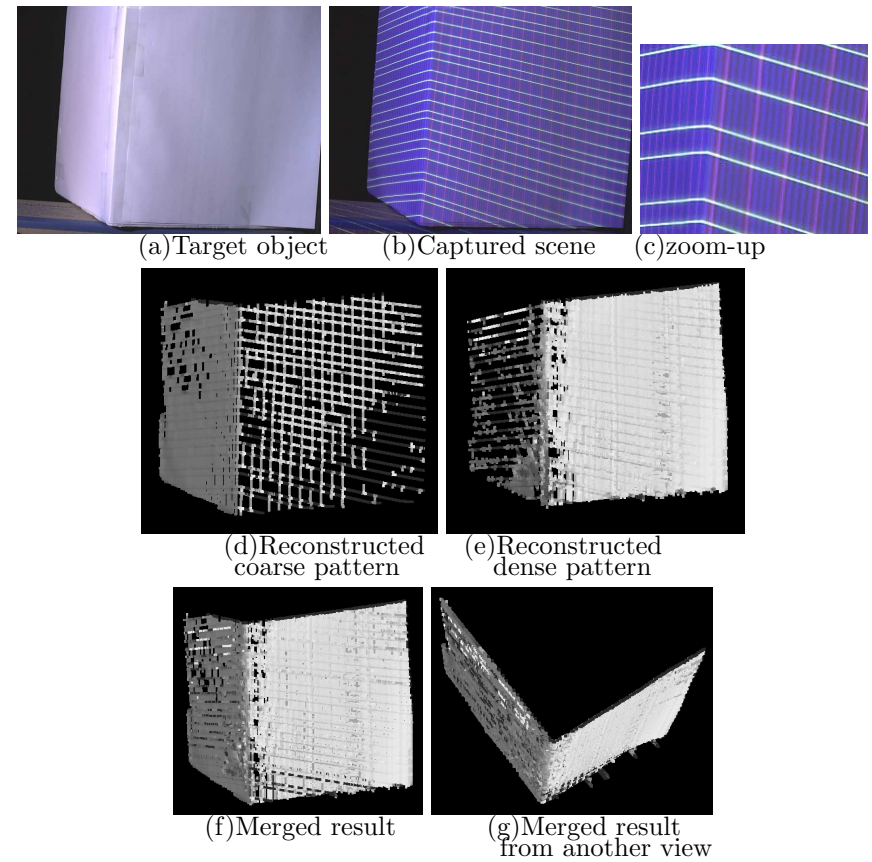
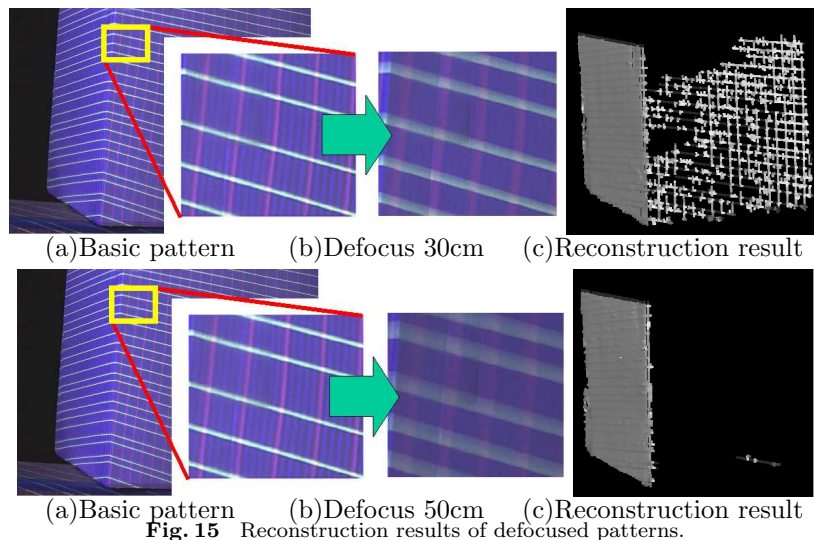


Fig. 14 Reconstruction results of coarse to fine technique: (a) captured scene, (b)(c) objects projected by the coarse-to-fine pattern, (d) reconstructed shape from coarse (sparse) pattern, (e) reconstructed shape from fine (dense) pattern, (f) and (g) merged results.

colors, the reconstruction succeeded. This shows that robust line detection was achieved even with a simple peak-detection algorithm, since only three colors (green for horizontal lines, red for coarse vertical lines, and blue for dense vertical lines) were used in the projected pattern.

In these experiments, curve detection was stable and no significant errors appeared if the scanning environment was not bright and the light intensity of the projector was high enough. However, if the environment was bright, the light intensity of the projector was low or the object was dark, curve detection became unstable and led to erroneous reconstruction. Using lasers or an infrared would be a simple solution.



5.5 Real data results

To demonstrate our technique, we scanned several types of objects with a real time system. We can reconstruct 3D shapes with an average 1.16 fps for a 1024×768 resolution camera. This is composed of 0.58 seconds for image capture, 0.22 seconds for reconstruction, and 0.07 seconds for rendering.

First, we scanned static objects which have abrupt depth and color changes.

Fig. 17 shows the captured scenes and results of reconstruction. In the experiment, a ceramic bottle, a paper mask and a ceramic jug with intricate shapes and textures were captured. As is apparent, detailed shapes were successfully recovered with the current method. Unlike previous one-shot methods^{(8),(10),(13),(15)}, the proposed technique achieves reconstruction even if jump edges and abrupt color changes exist so that lines are frequently segmented. In the results, we can see noticeable traces — as in Fig. 12, and for the same reason.

Next, dynamic scene reconstruction was conducted with a human face as the target object. The target scene was captured to obtain a series of images while the face was being moved and facial expressions changed freely. Fig. 18 shows sample captured scenes and the results of reconstruction. The proposed method successfully restored the human facial expressions with dense and accurate 3D point clouds. In the results, we can see noisy traces that are more significant than those in Fig. 12 and Fig. 17. We think that the main reason for the noise is because human faces have complicated reflectance properties such as non-uniform reflectance and subsurface scattering effects. Note that although there was a lot of noise, the overall shape is successfully reconstructed because the proposed method has a large tolerance for image processing errors (which is an advantage of the method).

Finally, to show the strength of a one-shot scanning method compared to a multi-frame scanning method, extremely fast motion was captured using a rotating fan as the target object. The target scene was captured with various shutter speeds as shown in Figs. 19 (b) to (d). In this case, only the fastest shutter speed (Figure 19 (d)) can capture the grid pattern with almost no motion blur, and we conducted the reconstruction using that image. Fig. 19 (e) shows the linked sets of detected curves of patterns, and Figs. 19 (f) to (h) show the results of reconstruction. The proposed method successfully restored the shape despite its extremely fast motion.

We also show the reconstruction result of the same fan with images affected by motion blur, captured with several different shutter speeds. Fig. 20 shows the results. In the results, as the exposure time becomes long, the image becomes more affected by motion blur and the reconstruction result becomes worse.

6. Conclusions

This paper proposes a technique to densely measure shapes of both static and dynamic objects using a single projection of grid pattern. The proposed technique does not involve encoding positional information into multiple pixels or color spaces, as often used in conventional ‘one-shot’ 3D measurement methods. Instead, the technique reconstructs the shape only from detected curves of a grid pattern. More specifically, coplanarity constraints between the detected grid points are used to obtain simultaneous equations, which can be solved up to a 1-DOF indeterminacy. The remaining ambiguity can be resolved by using known calibration data. The technique provides dense shape reconstruction even when discontinuities or occlusions in the shape are present. In addition, since it is necessary to distinguish only two types of patterns, such as vertical and horizontal curves, reconstruction is affected little by an object’s texture, providing robust shape reconstruction. Tests were carried out with simulated and real data, confirming that the proposed technique accurately and stably reconstructed 3D shapes with small computational cost.

Currently, simultaneous texture acquisition while continuously scanning dynamic scenes has not been implemented. Since the proposed method can be applied by using only two colors for a projected pattern, those colors can be easily replaced by infra-red spectra. Therefore, simultaneous texture acquisition may be possible in the future.

Acknowledgement

This work was supported in part by SCOPE No.072103013 (Ministry of Internal Affairs and Communications, Japan) and Grant-in-Aid for Scientific Research No.19700098 and 21700183 (Ministry of Education, Science, Sports and Culture, Japan).

References


- 1) J. Batlle, E. Mouaddib, and J. Salvi. Recent progress in coded structured light as a technique to solve the correspondence problem: a survey. *Pattern Recognition*, 31(7):963–982, 1998.
- 2) K.L. Boyer and A.C. Kak. Color-encoded structured light for rapid active ranging. *IEEE Trans. on PAMI*, 9(1):14–28, 1987.
- 3) D. Caspi, N. Kiryati, and J. Shamir. Range imaging with adaptive color structured light. *IEEE Trans. on PAMI*, 20(5):470–480, 1998.
- 4) J. Davis, D. Nehab, R. Ramamoorthi, and S. Rusinkiewicz. Spacetime stereo: A unifying framework for depth from triangulation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 27(2):296–302, Feb. 2005.
- 5) R. Furukawa and H. Kawasaki. Self-calibration of multiple laser planes for 3D scene reconstruction. In *3DPVT*, pages 200–207, 2006.
- 6) O. Hall-Holt and S. Rusinkiewicz. Stripe boundary codes for real-time structured-light range scanning of moving objects. In *ICCV*, volume 2, pages 359–366, 2001.
- 7) S. Inokuchi, K. Sato, and F. Matsuda. Range imaging system for 3-D object recognition. In *ICPR*, pages 806–808, 1984.
- 8) C. Je, S. W. Lee, and R.-H. Park. High-contrast color-stripe pattern for rapid structured-light range imaging. In *ECCV*, volume 1, pages 95–107, 2004.
- 9) H. Kawasaki and R. Furukawa. Shape reconstruction from cast shadows using coplanarities and metric constraints. In *ACCVLNCS 4843*, volume II, pages 847–857, 2007.
- 10) T.P. Koninckx and L.V. Gool. Real-time range acquisition by adaptive structured light. *IEEE Trans. on PAMI*, 28(3):432–445, March 2006.
- 11) T.P. Koninckx, T. Jaeggli, and L.V. Gool. Adaptive scanning for online 3d model acquisition. In *Sensor3D04*, pages 32–32, 2004.
- 12) D.J. Kriegman and P.N. Belhumeur. What shadows reveal about object structure. *Journal of the Optical Society of America*, 18(8):1804–1813, 2001.
- 13) J. Pan, P.S. Huang, and F.-P. Chiang. Color-coded binary fringe projection technique for 3-d shape measurement. *Optical Engineering*, 44(2):23606–23615, 2005.
- 14) M. Proesmans and L. Van Gool. A sensor that extracts both 3D shape and surface texture. In *Multisensor Fusion and Integration for Intelligent Systems*, 1996, pages 485–492, 1996.
- 15) J. Salvi, J. Batlle, and E.M. Mouaddib. A robust-coded pattern projection for dynamic 3D scene measurement. *Pattern Recognition*, 19(11):1055–1065, 1998.
- 16) K. Sato and S. Inokuchi. Range-imaging system utilizing nematic liquid crystal mask. In *Proc. of First ICCV*, pages 657–661, 1987.
- 17) J. Tajima and M. Iwakawa. 3-D data acquisition by rainbow range finder. In *ICPR*, pages 309–313, 1990.
- 18) T. Weise, B. Leibe, and L.V. Gool. Fast 3d scanning with automatic motion compensation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR’07)*, pages 1–8, June 2007.
- 19) M. Young, E. Beeson, J. Davis, S. Rusinkiewicz, and R. Ramamoorthi. Viewpoint-coded structured light. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2007.
- 20) L. Zhang, B. Curless, and S.M. Seitz. Spacetime stereo: Shape recovery for dynamic scenes. In *IEEE Computer Society Conference on Computer Vision and Pattern*

Recognition, pages 367–374, June 2003.

- 21) L.Zhang, N.Snaveley, B.Curless, and S.M. Seitz. Spacetime faces: High-resolution capture for modeling and animation. In ACM Annual Conference on Computer Graphics, pages 548–558, August 2004.

(Received November 28, 2007)

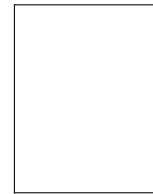
(Accepted July 17, 2007)



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 University in 2007. His research area includes shape-capturing, 3D modeling, appearance sampling and image-based rendering. He has won several academic awards including Songde Ma Outstanding Paper Award (best paper for ACCV) in 2007. He is a member of IPSJ, IEICE, and IEEE.



Hiroshi Kawasaki is an associate professor of Department of Information and Computer Sciences Faculty of Engineering, Saitama University, Japan. He received his ME in Information Engineering in 2000 and Ph.D. degree in Information and Communication Engineering from University of Tokyo, Japan, in 2003. He started working at Saitama University, Japan, in 2003. Prior to Saitama University, he worked at Microsoft Research Redmond, WA, USA, as an internship in 2000. He is also working as a visiting professor at INRIA Rhone-alpes, France, from 2009. His current research focus is on capturing a shape and texture and rendering them photo-realistically in the computer for VR/MR systems and ITS. He has published over 50 research papers including ICCV, CVPR, IJCV, 3DIM, Eurographics and MVA in computer vision, computer graphics and ITS and won several awards including Songde Ma Outstanding Paper Award (best paper for ACCV) in 2007. He is a member of IPSJ, IEICE, and IEEE.



Ryusuke Sagawa is an assistant professor at the Institute of Scientific and Industrial Research, Osaka University, Osaka, Japan. He received a BE in Information Science from Kyoto University, Kyoto, Japan, in 1998. He received a ME in Information Engineering in 2000 and Ph.D. in Information and Communication Engineering from the University of Tokyo, Tokyo, Japan in 2003. His primary research interests are computer vision, computer graphics and robotics (mainly geometrical modeling and visualization). He is a member of IPSJ, IEICE, RSJ, and IEEE.

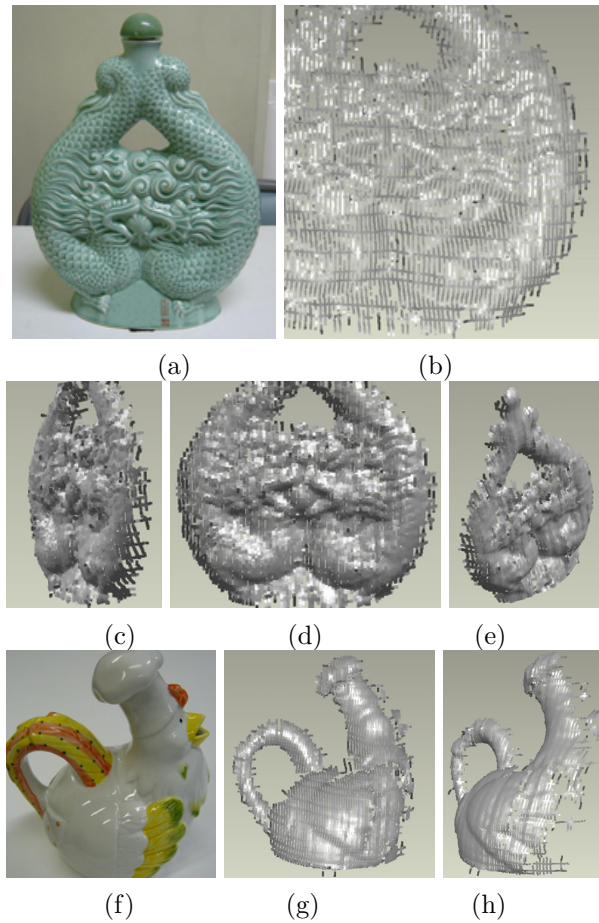


Fig. 17 Reconstruction results of static objects: (a)(f) the target objects, (b) close-up view of reconstructed shape of (a), and (c)(d)(e)(g) reconstructed shape.

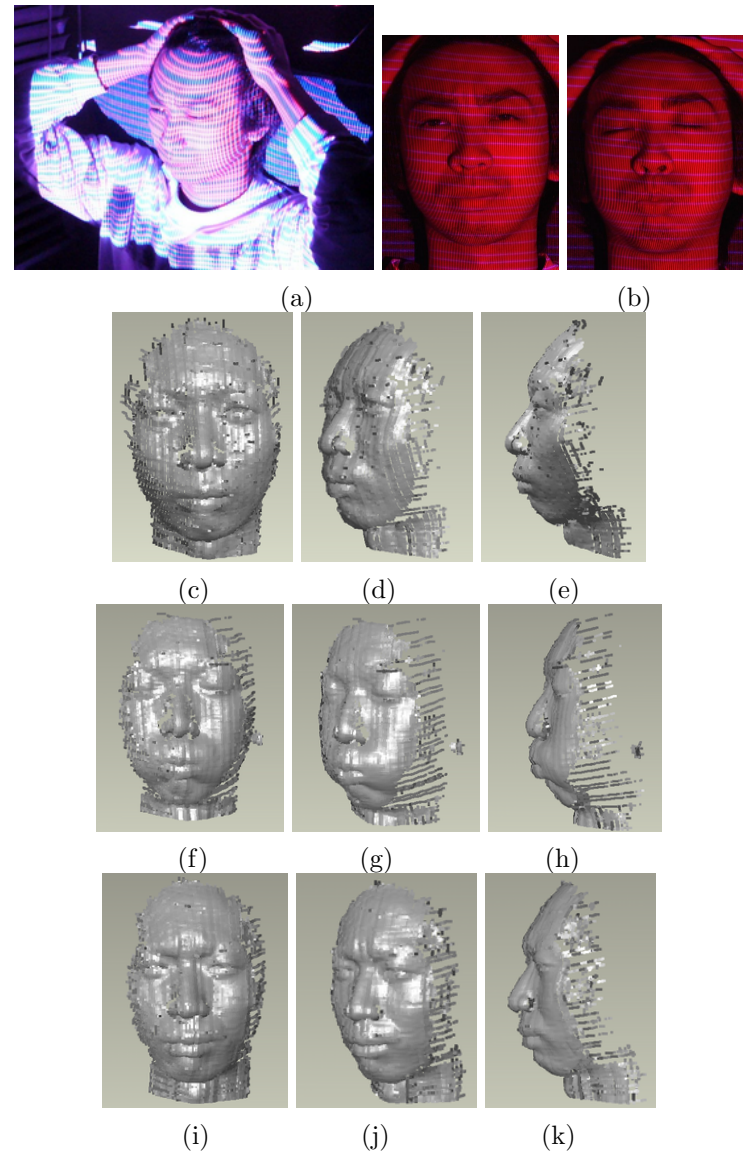


Fig. 18 Reconstruction of facial expressions: (a) a capturing scene, (b) the captured frames, (c)-(e) facial expression 1(normal), (f)-(h) facial expression 2(puffing up cheeks) and (i)-(k) facial expression 3(angry). Please carefully check the differences of the shape between the eyebrows of (d) and (j).

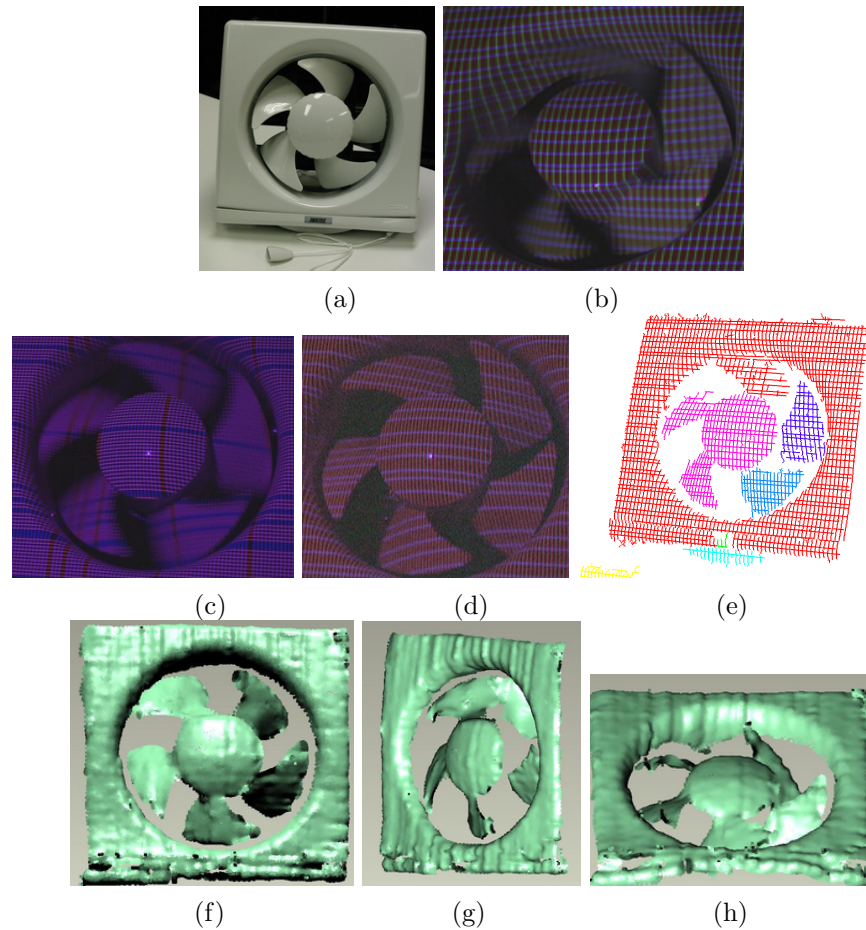


Fig. 19 Reconstruction of a fast rotating fan: (a) target object, (b) the captured frames with shutter speed 1/1000, (c) the captured frames with shutter speed 1/2400 (d) the captured frames with shutter speed 1/2550, (e) linked sets of detected curves (shown by colors), and (f) to (h) reconstructed shapes.

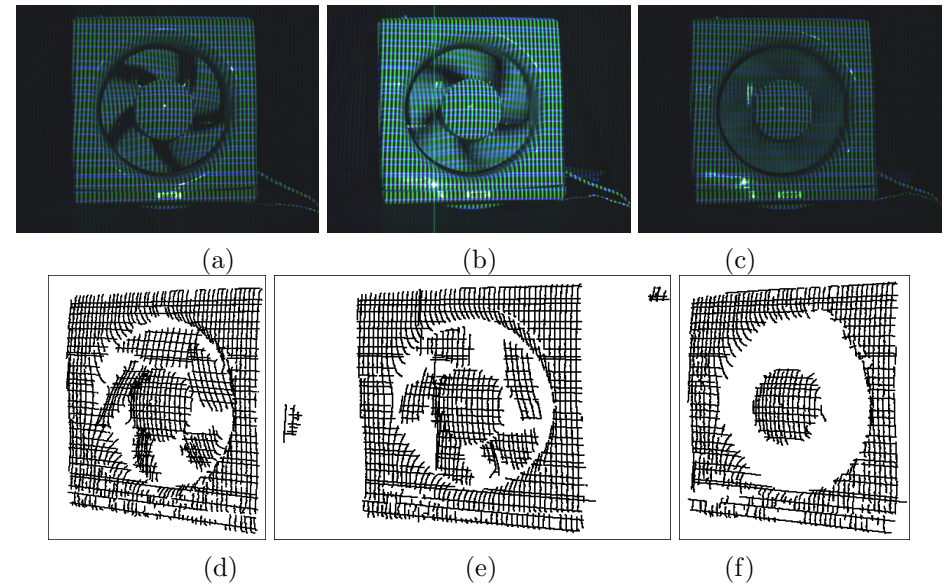


Fig. 20 Reconstruction of a rotating fan from images affected by motion blur: (a)-(c) captured image (shutter speeds are 1/2000, 1/800, and 1/250, respectively), (d)-(f) reconstruction results (shutter speeds are the same as (a)-(c)).