Dense 3D Reconstruction from High Frame-Rate Video using a Static Grid Pattern

Ryusuke Sagawa, Ryo Furukawa, and Hiroshi Kawasaki

***Abstract*—Dense 3D reconstruction of fast moving objects could contribute to various applications such as body structure analysis, accident avoidance, and so on. In this paper, we propose a technique based on a one-shot scanning method, which reconstructs 3D shapes for each frame of a high frame-rate video capturing the scenes projected by a static pattern. To avoid instability of image processing, we restrict the number of colors used in the pattern to less than two. The proposed technique comprises (1) an efficient algorithm to eliminate ambiguity of projected parallel-line patterns by using intersection points, (2) a batch reconstruction algorithm of multiple frames by using spatio-temporal constraints, and (3) an efficient detection method of color-encoded grid pattern based on de Bruijn sequence. In the experiments, the line detection algorithm worked effectively and the dense reconstruction algorithm produces accurate and robust results. We also show the improved results by using temporal constraints. Finally, the dense reconstructions of fast moving objects in a high frame-rate video are presented.**

***Index Terms*—Dense 3D Reconstruction, Projector-Camera Systems, Grid Patterns, Spatio-Temporal Analysis**

I. Introduction

Dense and precise shape acquisition of fast moving objects with high frame-rates has great potential for various fields, such as body structure analysis, accident avoidance and so on. For acquiring dense and accurate 3D shapes, many active 3D scanning systems using point-lasers or line-lasers have been developed. These methods essentially need time for scanning because point or line-lasers should be physically moved in one or two dimensions to scan a scene. Thus, they are not suitable for scanning fast-moving objects. Currently, 3D scanners that use area light sources such as video projectors to reduce scanning times are actively researched [1], [2]. Among them, spatial encoding methods that use only a single input image are considered to be suitable for capturing fast moving objects. A typical spatial encoding method uses a color code, where reconstruction is achieved by stereo with a window matching technique [1], [3]. However, since color coding methods are easily affected by textures and shapes of the objects, they originally have critical problems for density and precision on shape reconstruction.

In this paper, two approaches are proposed to solve the problem by using grid patterns. As the first approach, we propose a unified formulation for grid-based 3D reconstruction with single [4] and multiple [5] projectors based on coplanarity constraints. For the second approach, we propose a new formulation of the problem that can efficiently utilize the properties of the grid patterns based on epipolar constraint to achieve stable and dense reconstruction.

The first approach relates to shape reconstruction techniques that use intersections of line patterns [6], [7], [8]. Although these techniques have advantages that they require no identification of line patterns, those methods are based on solving a large problem of linear equations, which is not efficient enough for real-time systems. In addition, these techniques are two step algorithms and require encoding of the patterns into intervals of lines which leads to sparse reconstruction. In the second approach, we re-formulate the problem by an efficient calculation technique that can drastically reduce the number of variables by utilizing adjacency information between grid lines. Moreover, if the input is an image sequence from videos, temporal constraints between frames can be utilized to improve the results. In this paper, batch reconstruction algorithm for multiple frames using spatio-temporal constraints is proposed.

Another contribution of the paper is a dense grid pattern using de Bruijn sequence. Normally, the number of required colors for de Bruijn sequence increases if unique identifications for all the pixels are necessary. In the proposed method, however, a small number of colors are sufficient, because non-unique identifications with the same IDs appearing periodically are allowed for restoring the scene with our reconstruction method. Moreover, the grid pattern can be dense because errors of the identification are corrected up to some extent with our reconstruction method. To achieve stable extraction of the grid pattern, connectivity of vertical and horizontal pixels are efficiently utilized by a belief propagation technique.

II. Related works

Shape reconstruction techniques with a structured light using temporal and spatial coding are summarized in [9]. Systems using only temporal coding are easy to be implemented and have good features such as accuracy, denseness and robustness, therefore they have been used for many real applications [10], [11], [12]. Since the technique needs to capture multiple images, it is not suitable for high-speed capturing. Recently, several methods for high-speed capturing were proposed by using a DLP projector and a high-speed camera [13], [14], [15]. Zhang *et al.* [13] and Weise *et al.* [14] proposed a stereo system with active illumination based

R. Sagawa is with the Intelligent Systems Research Institute, National Institute of Advanced Industrial Science and Technology, Ibaraki, Japan. e-mail: ryusuke.sagawa@aist.go.jp .

R. Furukawa is with the Graduate School of Information Science, Hiroshima City University, Hiroshima, Japan. e-mail: ryo-f@hiroshima-cu.ac.jp .

H. Kawasaki is with the Department of Information and Biomedical Engineering, Kagoshima University, Kagoshima, Japan. e-mail: kawasaki@ibe.kagoshima-u.ac.jp .

on phase-shift. Narasimhan *et al.* [15] proposed a method to recognize high-speed temporal codes produced by a DLP projector. Though these methods can capture range images at high frame rate, the motion of objects must be relatively slow in the image sequence to recognize the temporal codes. Also, in several researches, the required number of patterns are reduced using both temporal and spatial changes [2], [16]. Hall-Holt *et al.* proposed an enhanced method to eliminate the limitation by aligning the reconstructed shape with an assumption of a limited maximum speed [2]. Davis *et al.* proposed an efficient method to reduce patterns by using multiple cameras [16]. However, since the techniques have several restrictions as stated above, they are not suited for capturing objects with fast motion. Some methods [1], [3] use projectors only to capture textures that change over time and restore 3D information in similar manner with passive stereo techniques. Since they still require several patterns for identification, they are not suited for extremely fast moving objects.

Techniques using only spatial encoding of a pattern are suitable for capturing fast-moving objects, since they use only a single image [17], [18], [19], [20], [21]. On the other hand, there are other problems, such as they typically need complex patterns or colors to encode positional information of the pixel of the projector. To determine the spatial codes uniquely, the size of the code becomes large. Such patterns are easily affected by textures, shape discontinuities and pattern compression caused by tilted surfaces. In addition, decoding of individual points is done with neighboring points, so the pattern becomes more complex and image processing is usually difficult. Thus, 3D reconstruction tends to be sparse and unstable. Kinect [22] is one of the recent successful product of 3D scanner with well-balanced accuracy, resolution and cost effectiveness. It is a projector-camera system that reconstructs 3D shapes from a single image. Although Kinect can be efficiently used for the purpose of motion capture and gesture recognition, 3D quality is not enough for modeling and inspection purposes.

There are several researches that achieve shape reconstruction from single images composed of multiple lines or stripes. Koninckx *et al.* proposed a technique for dense shape reconstruction using a simple pattern, *i.e.*, a set of stripes [23]. Similarly, Frueh and Zakhor [24] used dense vertical stripes and sparse horizontal lines. The vertical stripes are used for shape reconstruction and the horizontal lines for identifying the vertical 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. In our method, since it detects discontinuities by using color codes, the shape is correctly reconstructed even if discontinuities exist. In some techniques, 3D scenes are reconstructed by regarding projected line patterns as intersections between 3D planes and the scenes [6], [7], [8], however, their technique can be only applied for sparse patterns. To avoid erroneous matching in finding correspondences caused by self-similarities of uniformly-spaced grid patterns, Ulusoy *et al.* proposed an irregularly-spaced grid pattern based on de Bruijn sequence [25], while the method proposed in this paper introduces a color code based on de Bruijn sequence.

III. 3D RECONSTRUCTION FROM A STATIC GRID PATTERN

A. Overview

We propose methods to reconstruct the shape from a grid pattern which consists of two directional parallel lines. Since all the lines of parallel line sets are identical, correspondences between the observed projected curves and the lines on the original pattern cannot be determined by just observing the lines themselves. Such ambiguity is efficiently solved by using either coplanarity constraints [7], [8] or epipolar constraints [26], and shapes are consequently reconstructed.

To project such multiple parallel line sets, using either single projector as shown in Fig.1(a) or multiple projectors as shown in Fig.1(c) can be considered. With our method, it is assumed that the camera and the projector are calibrated, *i.e.*, the intrinsic parameters of the devices and their relative positions and orientations are known. Since the projected pattern is fixed and does not change during the entire process of scanning, no synchronization is required. An example of the projected patterns for Fig.1(a) and (c) are shown in Fig.1(b) and (d), respectively. In the case of multiple projectors, each projector is required to cast just single directional parallel line sets, and the grid patterns are formed on the surfaces of the target objects [5]. In this paper, we explain the solution to obtain the 3D shape by the unified approach of the both setups, followed by a novel extended method using temporal consistency which realizes more robust and accurate reconstruction.

The flow of the proposed methods is summarized in Fig.2. First, dense grid patterns are detected in each frame. To detect them stably, a method which uses belief propagation is proposed. Second, the curve ID of each curve which is detected in the camera image is determined by using the connectivity of curves. In the case of batch reconstruction of multiple frames, the correspondences of curves between frames are also detected and the same ID on the previous frame is assigned. The details of those image processing are described in Sec.IV A-C. Finally, 3D shapes are reconstructed from the positions of the detected grid points and the IDs of the curves. The reconstruction methods by using coplanarity constraints and epipolar constraints are explained in Sec.III B and C, respectively. Then spatio-temporal extension of the method is explained in Sec.III D. Once the 3D shape of the grid pattern is obtained, dense 3D shape is reconstructed by interpolating the pixels between the lines; the details are described in Sec.IV D.

B. Shape from grid pattern based on coplanarity constraint

In this section, 3D reconstruction based on coplanarity constraints is explained. A set of patterns of parallel lines is used as a structured light. Each of the line patterns emitted from the projector sweeps a plane in the 3D space. We use a term *light plane* to represent such a plane. Then, lines projected onto the target scene are detected from the image. Certainly each detected line corresponds to a light plane;

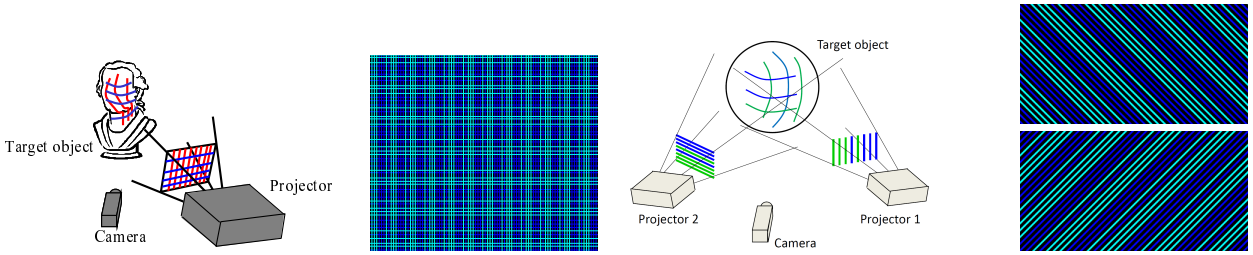


Fig. 1. (a) Scanning system with one projector: multiple lines are projected and their intersections are detected and used for reconstruction. (b) a projected pattern for the system (a). (c) another setup of the proposed system with two projectors. (d) a projected pattern for the system (c).

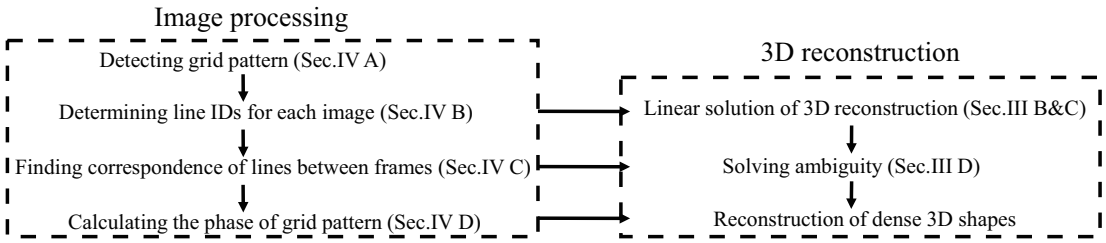


Fig. 2. The algorithm of the proposed method: first, dense grid patterns are detected in each frame. Second, the curve ID of each curve is tentatively determined for each frame. Third, the correspondences of curves between frames are found. After calculating the phase of grid patterns for every pixel, the proposed method generates the dense 3D shapes.

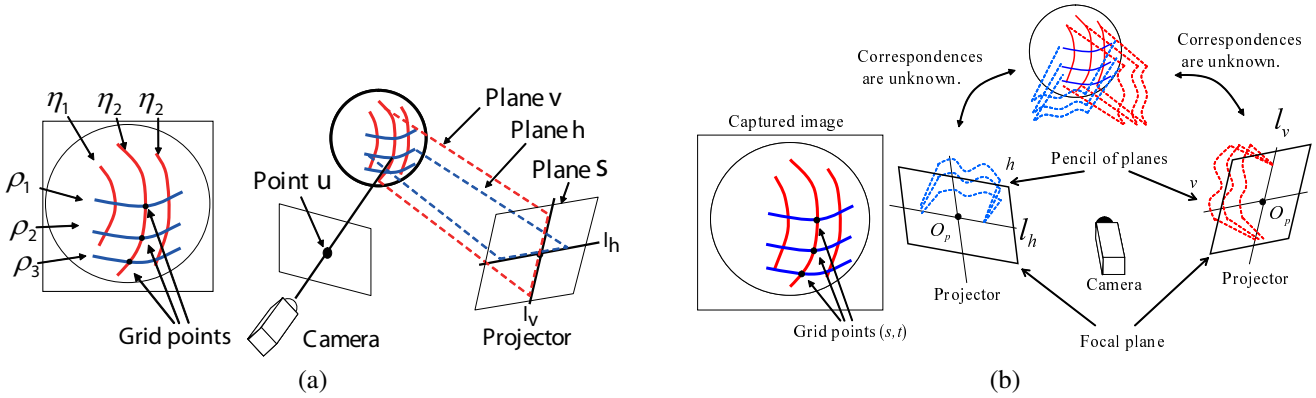


Fig. 3. A light plane is formed by a grid line in the projected pattern. (a) and (b) are the cases of one and two projectors. Parallel lines in the pattern share an axis of the pencil of planes. l_V and l_h are the direction vectors of the axes for vertical and horizontal lines.

however, the actual position of the plane cannot be determined just by observing the lines.

Fig.3 shows that each line of the grid pattern forms a light plane in the 3D space and is observed as a curve in the camera image. Fig.3 (a) and (b) are the cases with one and two projectors, respectively. In both cases, the grid pattern is formed by horizontal and vertical lines, which sweep horizontal and vertical light planes in the 3D space. We assume both the vertical and horizontal planes do not include the optical center of the camera. Thus, the vertical plane can be represented by $\mathbf{v} \cdot \mathbf{x} + 1 = 0$, where $\mathbf{v}$ is a 3D vector of the parameters of the plane and $\mathbf{x}$ is a 3D point in the camera coordinates.

All the vertical planes include the axis of the pencil of planes. Let the 3D line of the axis be represented as $\mathbf{p}_v + t\mathbf{l}_\mathbf{v}$, where $\mathbf{p}_v$ is a point and $\mathbf{l}_\mathbf{v}$ is the directional vector. Then, $\mathbf{v}$ should fulfill $\mathbf{v} \cdot \mathbf{p}_v + 1 = 0$ and $\mathbf{v} \cdot \mathbf{l}_\mathbf{v} = 0$. By solving these equations, we obtain

$$\mathbf{v} = \mathbf{v_0} + \eta(\mathbf{p}_v \times \mathbf{l}_\mathbf{v}) = \mathbf{v_0} + \eta\mathbf{v}', \tag{1}$$

where $\mathbf{v_0}$ is an arbitrary vertical plane, and $\mathbf{v}' \equiv \mathbf{p}_v \times \mathbf{l}_\mathbf{v}$. Similarly, a horizontal plane $\mathbf{h}$ can be represented by $\mathbf{h} = \mathbf{h_0} + \rho(\mathbf{p}_h \times \mathbf{l_h}) = \mathbf{h_0} + \rho\mathbf{h}'$, where $\mathbf{p}_h + t\mathbf{l_h}$ is the axis of the pencil of planes, $\mathbf{h_0}$ is an arbitrary horizontal plane, and $\mathbf{h}' \equiv \mathbf{p}_h \times \mathbf{l_h}$.

If a grid point (*i.e.*, an intersection point between vertical and horizontal curves, some of which are shown in Fig.3) is detected, the vertical and horizontal curves are included by a vertical plane $\mathbf{v}$ and horizontal plane $\mathbf{h}$, respectively. Therefore, the following equation is fulfilled:

$$\mathbf{u} \cdot (\mathbf{v} - \mathbf{h}) = 0, \tag{2}$$

where $\mathbf{u} \equiv (u, v, 1)$ represents the 2D position of the detected grid point (u, v) in the normalized camera coordinates (see

Fig.3(a)). Intuitive meanings of this equation can be found in the work [7]. From the above equations, we obtain

$$(\mathbf{u}\cdot\mathbf{v}')\eta-(\mathbf{u}\cdot\mathbf{h}')\rho=-\mathbf{u}\cdot(\mathbf{v_0}-\mathbf{h_0}). \tag{3}$$

This is a linear equation with respect to η and ρ, and a constraint to determine the parameters of planes, which we call a *coplanarity constraint*. It is common for the both cases of one and multiple projectors shown in Fig.3(a) and (b).

By accumulating Eq.(3) for all the grid points, a linear equation

$$\mathbf{Ax}=\mathbf{b} \tag{4}$$

is obtained, where $\mathbf{A}$ is the coefficient matrix, $\mathbf{x}$ is the parameter vector $[\eta_1\dots\eta_M\rho_1\dots\rho_N]^\top$, $\mathbf{b}$ is the vector of the constant terms, and M and N are the numbers of the detected vertical and horizontal curves. If this linear equation is solved, the 3D shapes of the detected curves are calculated by triangulation between the light planes and the lines of sights.

In the case of one projector, the linear solution is not unique and has one degree of freedom. In this case, the two pencils of planes share one plane (plane S in Fig.3(a)). By setting this plane to be $\mathbf{v_0}=\mathbf{h_0}$, Eq.(4) becomes $\mathbf{Ax}=\mathbf{0}$. The solution is obtained by calculating the eigenvector associated with the minimum eigenvalue of $\mathbf{A}^\top\mathbf{A}$. The 1-DOF indeterminacy of the equations pointed out by Furukawa [4] corresponds to the scaling ambiguity of the eigenvector.

The 1-DOF indeterminacy can be removed by using knowledge of the projected pattern. Since the system is calibrated, the actual patterns cast by the projector are known. Thus, the solution can be acquired by matching the solution with the known pattern using 1D search [4]. For this reason, the pattern should be locally unique and the line intervals can be used as a simple solution, *i.e.*, random intervals [4]. Additionally, since the projected pattern in this paper is encoded by a periodic color code, this constraint contributes to improve the robustness of matching [27].

The method with two projectors resolved this ambiguity by setting the projectors so that their axes $\mathbf{l_v}$ and $\mathbf{l_h}$ are skew (*i.e.* do not intersect in 3D space). In this case, the constant vector $\mathbf{b}$ becomes nonzero because the plane shared by the two pencils of planes does not exist. The solution is uniquely calculated by $\mathbf{x}=(\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top\mathbf{b}$ as proven in [28]. Since the 1D search is not necessary for the case of two projectors, both vertical and horizontal lines can be dense with uniform intervals, which improves the reconstruction of thin and small objects [28].

C. Unique linear solution based on relative coordinate and epipolar constraint

In this paper, we propose another approach to obtain a unique solution by the unified formulation for the cases of both one and multiple projectors. In the above formulation, the line intervals are only used for matching between patterns after the coplanarity solution. In the new approach, we use the intervals as an additional constraint. Especially if the interval of the lines is uniform, the parameter becomes $\eta_{i+1}-\eta_i=const_v$, where η_{i+1} and η_i are the parameters of adjacent curves in a linked set of curves, which share the intersection points. Similarly, the constraint for horizontal curves is defined as $\rho_{j+1}-\rho_j=const_h$. These constraints can be directly added to Eq.(4). Once this constraint is introduced, the problem of 3D reconstruction from grid patterns can be explained based on epipolar constraints.

First, we consider the case of one projector. A pixel (x_c,y_c) is an intersection point of a vertical curve and a horizontal curve. Since the projected lines are encoded by a periodic color code, each curve has a periodic ID. In Fig.4(a), the vertical curve is k_v-th curve in m_v-th cycle of the periodic pattern in the linked set of curves, where c is the number of lines in a cycle. Similarly, the horizontal curve is k_h-th one in m_h-th cycle. The coordinate of the intersection point in the projector coordinate system is represented by $(x_p,y_p)=(w_v((m_v+s)c+k_v),w_h((m_h+t)c+k_h))$, where w_v and w_h are the intervals between lines. s and t are the offsets of cycles, which are unknown from the curve detection since the coordinate calculated by curve detection is relative one in a linked set of curves.

This correspondence gives the following epipolar constraint:

$$[x_p,y_p,1]\mathbf{F}[x_c,y_c,1]^\top=0, \tag{5}$$

where $\mathbf{F}$ is the fundamental matrix between the camera and the projector. Since the unknown variables in this equation are only s and t, Eq.(5) is a linear equation, which is obtained for all the intersection points. The parameters are shared by the pixels in the same linked set. If the numbers of linked sets of vertical and horizontal curves are N_s and N_t, the following simultaneous linear equation is obtained:

$$\mathbf{Bx}=\mathbf{b},\quad \mathbf{x}=[s_1,\dots,s_{N_s},t_1,\dots,t_{N_t}]^\top, \tag{6}$$

where $\mathbf{B}$ is the coefficient matrix and $\mathbf{b}$ is the vector of constant terms. The number of variables is N_s+N_t, which is much smaller than the number of curves in the camera image.

The second setup is the case of multiple projectors. If the points of projector 1 and 2 corresponding to a camera point (x_c,y_c) are $(x_{p1},y_{p1})=(w_1((m_1+s)c+k_1),s')$ and $(x_{p2},y_{p2})=(w_2((m_2+t)c+k_2),t')$, respectively, the epipolar constraints are expressed as follows:

$$[x_{p1},s',1]\mathbf{F}_1[x_c,y_c,1]^\top=0 \tag{7}$$

$$[x_{p2},t',1]\mathbf{F}_2[x_c,y_c,1]^\top=0 \tag{8}$$

$$[x_{p1},s',1]\mathbf{F}_{12}[x_{p2},t',1]^\top=0, \tag{9}$$

where $\mathbf{F}_1$, $\mathbf{F}_2$ and $\mathbf{F}_{12}$ are the fundamental matrices between the camera, projector 1 and 2. In this formulation, each projector emits a set of lines parallel to the vertical axis of the projector image. If the direction of lines is different, the representation of projector points can be modified without loss of generality. This situation is illustrated in Fig.4(b). The corresponding points are defined by the three pairs of epipolar lines. s' and t' are auxiliary variables, which are expressed by linear polynomials of s and t from Eq.(7) and (8), respectively. By substituting s' and t', Eq.(9) is expressed by the terms of st,s,t, and constant.

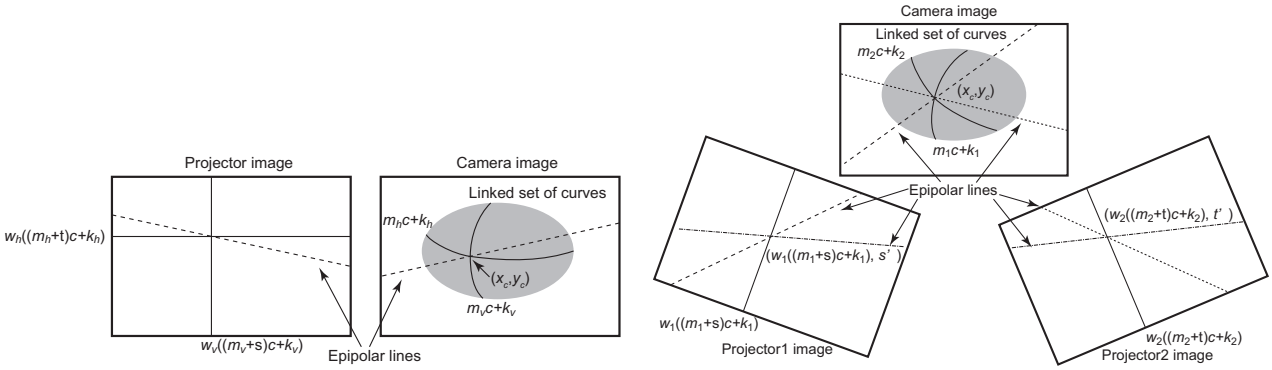


Fig. 4. (a) The pixel (x_c, y_c) is an intersection point in a linked set of curves. The vertical and horizontal curves are k_v-th curve in m_v-th cycle and k_h-th one in m_h-th cycle of the periodic pattern in the linked set of curves, respectively. (b) The relationships between three corresponding points are given by the epipolar geometry. (x_c, y_c) is corresponding to $(w_1((m_1 + s)c + k_1), s')$ and $(w_2((m_2 + t)c + k_2), t')$, respectively. s' and t' are expressed by linear polynomials of s and t.

To make Eq.(9) linear, we introduce a new variable $r = st$ to replace the second-order term. Eq.(9) can be regarded as a linear equation of r, s and t. Similar to Eq.(6), r, s and t can be solved by gathering the constraints of all the intersection points. If $\mathbf{B}$ in Eq.(6) is full rank, a unique solution is obtained. Since each row of $\mathbf{B}$ means the epipolar line at the intersection point in the projector coordinate system, $\mathbf{B}$ is full rank if they are not parallel.

The proposed formulation can be extended to the case of three or more projectors. If m projectors are used, a pair of patterns is chosen out of them and the constraint equations are created by using the regions of the pair. The simultaneous equation is obtained by using $\binom{m}{2}$ combinations of patterns.

Since the nonlinear constraint $r = st$ is omitted from the system in the above linear solution, the obtained solution may not satisfy it in some cases. Especially, it happens when the regions of three or more patterns are overlapped. Therefore, we use the linear solution as an initial guess and improve it by nonlinear minimization of Eq.(9).

Since s and t are the numbers of cycles, they are integer values. The linear solution of Eq.(6), however, can be non-integer due to the error of calibration and image processing. The problem, therefore, becomes an integer least square (ILS) problem, in which the linear equation is solved under the constraint that the solution must be a set of integers. It is known that this problem occurs in the case of GPS measurement [29]. In this paper, we implemented an ILS solver based on one of the solvers, MILES [30], to obtain s and t.

D. Batch reconstruction of multiple frames with spatio-temporal constraint

The method explained heretofore reconstructs 3D shapes frame by frame. If the motion of objects is not so fast compared to the frame rate of the video, the correspondence between frames can be found by image processing, which can improve the robustness of reconstruction as the spatio-temporal constraint.

Now, we assume that the corresponding points between frames are given by the method described in Sec.IV-C. If two points at frame T and $T-1$ have the correspondence, their corresponding points in the projector coordinate are represented by $(w_v((m_{v,T} + s_T)c + k_{v,T}), w_h((m_{h,T} + t_T)c + k_{h,T}))$ and $(w_v((m_{v,T-1} + s_{T-1})c + k_{v,T-1}), w_h((m_{h,T-1} + t_{T-1})c + k_{h,T-1}))$, respectively. Since their periodic curve IDs are the same, $k_{v,T} = k_{v,T-1}$ and $k_{h,T} = k_{h,T-1}$. The parameters at frame T become

$$s_T = s_{T-1} + m_{v,T-1} - m_{v,T}, \tag{10}$$

$$t_T = t_{T-1} + m_{h,T-1} - m_{h,T}. \tag{11}$$

It indicates that the parameters at frame T can be replaced by those at frame $T - 1$, and thus, the constraints can be extended to multiple frames. For the frames $f = 1 \dots T$, the parameters at frame f are calculated by solving $\mathbf{B}_f \mathbf{x}_f = \mathbf{b}_f, \mathbf{x}_f = [s_{1,f}, \dots, s_{N_s,f}, t_{1,f}, \dots, t_{N_t,f}]^\top$ independently. If all the parameters at frames $2 \dots T$ have correspondences between adjacent frames, it can be modified as

$$[\mathbf{B}_1 \cdots \mathbf{B}_T]^\top \mathbf{x}_1 = [\mathbf{b}_1, \cdots \mathbf{b}_T]^\top. \tag{12}$$

The parameters at frame 1 are estimated by using all the frames, and the rest of parameters are calculated from $\mathbf{x}_1$. Since the number of parameters in Eq.(12) is much smaller than that of frame-by-frame solution, the robustness can be improved.

E. Discussions about grid-based 3D reconstructions

To conclude this section, we discuss the relationships between two 3D reconstruction approaches described in this section. One approach uses coplanarities of the grid points (Sec. III-B). With this approach, it was revealed that the solution of coplanarity constraints has 1-DOF indeterminacy [8]. Such ambiguity can be efficiently solved by using positional information of the projected lines that are predetermined and known. Since the combination of the coplanarity constraints and the positional information of the projected lines is equivalent to epipolar constraints, it can be understood that the coplanarity-based approach also implicitly uses epipolar constraints.

On the other hand, the second approach explicitly uses the epipolar constraints to solve the ambiguity (Sec. III-C). In this method, lines are arranged to configure the periodic pattern and the ambiguity of the phase is efficiently solved; Note that the problem is same as the phase unwrapping process of the phase-shifting methods [31], [13], [14]. Several approaches have been proposed to solve the problem with a single image, e.g. a composite pattern of multiple phases [32], and the assumption of small motion between frames [33]. Therefore, our solution is an alternative solution for phase unwrapping problem using geometric information (epipolar constraint in this case) instead of using several different phases in previous methods. One advantage of our method compared to the previous methods is that our method requires only two patterns; such a feature is critical to one-shot scan.

Pros and cons of the two approaches is as follows. Since the coplanarity method does not assume the adjacency relationship on detected lines, shapes can be reconstructed even if the lines are poorly detected and/or some lines are missing. However, it remains one degree of indeterminacy and should be solved by additional matching process, resulting in sparse reconstruction and high computational cost. On the other hand, with the approach using epipolar constraints, although robust line detection is required to ensure the adjacency between lines, the solution can be directly retrieved by solving linear constraint equations without any other information; with this feature, dense reconstruction is realized by using a dense regular grid pattern.

IV. Image processing with grid patterns

The detection of color line patterns proposed in this paper consists of three steps. The first step is detecting curves and discriminating them for vertical and horizontal directions regardless of the colors, and the second step is decoding the periodic color code based on the de Bruijn pattern as shown in Fig.1 (b). The third step is calculating dense phase of periodic pattern by interpolating curves.

In the previous studies using stripe patterns, many colors were used to get unique correspondences. In that case, the cross talk of RGB spectrum severely affected the image processing. Since our method uses maximum two colors, we can detect the line as a peak of intensity regardless of color information in the perpendicular direction of the line, which can efficiently avoid the cross talk problem. By using our method, dense lines that exist every other pixel can be detected in theory.

A. Detecting lines of grid pattern

The pattern emitted from one or multiple projectors consists of multiple sets of parallel lines. First, we detect a set of curves in a camera image as the projection of parallel lines by discriminating them from the other sets of curves. We use the direction and color of curves for discrimination. We assume that curves in a camera image to be detected are nearly vertical curves, because we can rotate the image if the directions of the target curves are different. Additionally, the colors of lines are encoded by using two of RGB colors, for example blue and cyan, in this paper. In this case, all lines have blue component and curves are detected by using blue plane of the camera image.

A curve is regarded as a set of connected pixels that has a peak value of intensity along the direction perpendicular to the curve. In this paper, instead of detecting peak pixels directly, we consider the problem as the segmentation of pixels with respect to the derivative of the intensity along an axis. Consequently, a curve is detected as a boundary of different segments.

To take the continuity of a curve into consideration, we propose a method of segmentation based on the belief propagation (BP) [34], which is an energy minimization problem on a graph defined by

$$E(f) = \sum_{p \in V} D_p(f_p) + \sum_{(p,q) \in U} W_{pq}(f_p, f_q), \qquad (13)$$

where f is the set of labels to be determined, V is the set of nodes, U is the set of edges, and p and q are the nodes of the graph. $D_p(f_p)$ is the data cost of assigning label f_p to p. $W_{pq}(f_p, f_q)$ is the discontinuity cost of assigning labels f_p and f_q to neighboring nodes. Graph cut [35] can be an alternative as a method to solve an energy minimization problem, but we chose BP because the cost computed in BP is also used to detect curve position in sub-pixel accuracy as described later.

In detecting curves, the nodes correspond to pixels of the camera, and the edges are the connection to 4-neighboring pixels. The proposed method discriminates every pixel to three labels based on the derivative of the intensity along x-axis (horizontal) of the camera. The labels are positive (P), negative (N), and nearly zero (0). From the definition of the curve defined above, the curve is detected as the boundary of the labels P and N.

The data cost $D_p(f_p)$ is computed by the forward difference

$$D_p(f_p) = \begin{cases} I(x+1,y) - I(x,y) & \text{if } f_p = N \\ |I(x+1,y) - I(x,y)| & \text{if } f_p = 0 \\ -(I(x+1,y) - I(x,y)) & \text{if } f_p = P \end{cases}, \qquad (14)$$

where $I(x,y)$ is the intensity at a pixel $p = (x,y)$. The discontinuity cost $W_{pq}(f_p, f_q)$ depends on the direction of the edge as follows:

$$W_{pq}(f_p, f_q) = \begin{cases} -\lambda(f_q - f_p)(I(x+1,Y) - I(x,y)) \\ \quad \text{if the edge } (p,q) \text{ is along x-axis} \\ |f_q - f_p| \\ \quad \text{if the edge } (p,q) \text{ is along y-axis} \end{cases},$$

where f_p and f_q are 0, 1, and 2 for the labels N, 0, and P, respectively. λ is a user-defined parameter. Because of changing the discontinuity cost with respect to the direction of the edges, the proposed method can detect vertical curves while the horizontal curves are ignored even if they are the same color. The label of each node is determined by choosing the label of the minimum cost. If the boundary of the labels from P to N is detected, the subpixel x-coordinate of a point in the curve is computed by

$$x + \frac{C_p(N) - C_p(P)}{(C_p(N) - C_p(P)) - (C_q(N) - C_q(P))}, \qquad (15)$$

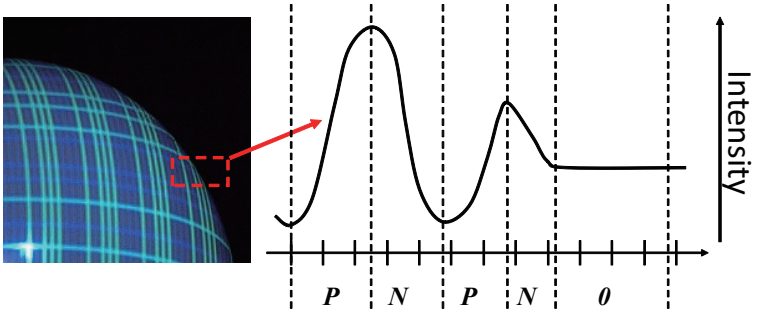


Fig. 5. The proposed method discriminates every pixel to three labels based on the derivative of the intensity along x-axis (horizontal) of the camera. The labels are positive (P), negative (N), and nearly zero (0). This figure shows an example of vertical curves.

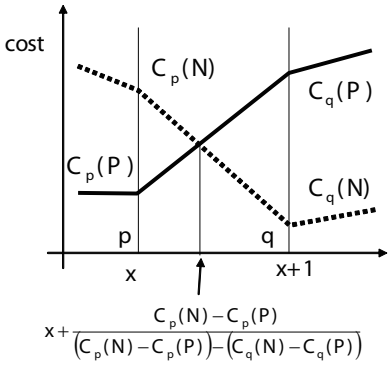


Fig. 6. The subpixel position of a curve is computed by interpolating the crossing position of the costs.

where $C_p(f_p)$ is the final cost, which is computed as the sum of the data cost and the messages (Fig. 6). The x-coordinates of p and q are x and $x+1$, respectively. A curve is given by connecting detected points along vertical direction. Since the effects of horizontal curves are removed in the cost when detecting vertical curves, we interpolate the final cost instead of using the input image directly.

Fig. 7 shows the example of detecting vertical curves. (a) is the input image. (b) and (c) shows the changes of labels by message passing. Each pixel classified to the labels P (red), N (blue), and 0 (black). The effect of horizontal curves remains in the initial state and it disappears after ten iterations.

B. Determining periodic color code

Next, we decode the color code based on the de Bruijn sequence [17], [20], [36]. 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 presented exactly once. If a projected pattern is encoded by two or more symbols distinguished in a camera image, 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.

There are three advantages to use color patterns. First, wrong connection of curves can be reduced by color information. The second advantage is that matching the sets of planes becomes easy by using color patterns as described in Sec. III-B. Without color patterns, all combinations between projected and observed patterns must be compared to solve the ambiguity. The third advantage is that it is useful for finding the correspondence of grid points between frames when the shapes of multiple frames are simultaneously reconstructed based on the method described in Sec.III-D. The IDs assigned to vertical and horizontal curves can be used as the feature for matching points. The probability of wrong correspondence drastically decreased by using the IDs.

For robust decoding, two dimensional regularization based on BP is used in our method. We use the grid points as the node of a graph and the edges are determined by the detected curves as shown in Fig. 8. Since each cycle consists of eight lines, the number of IDs is eight. The data cost $D_p(f_p)$ for decoding the de Bruijn sequence in the vertical curves is defined as follows:

$$D_p(f_p) = |H(p) - H(f_p)|, \tag{16}$$

where $H(p)$ is the hue of the grid point, and $H(f_p)$ is the hue of the projected light for the curve of $f_p = 0, \ldots, 7$. In actual case, since the color of a grid point is affected by both vertical and horizontal curves, $H(p)$ is computed as the hue at the midpoint between the grid point p and its neighbor grid points. Fig. 8 shows an example of determining the hue of a vertical curve that consists a grid point. It is computed as the average of two midpoints p_1 and p_2.

The discontinuity cost $W_{pq}(f_p, f_q)$ is given by

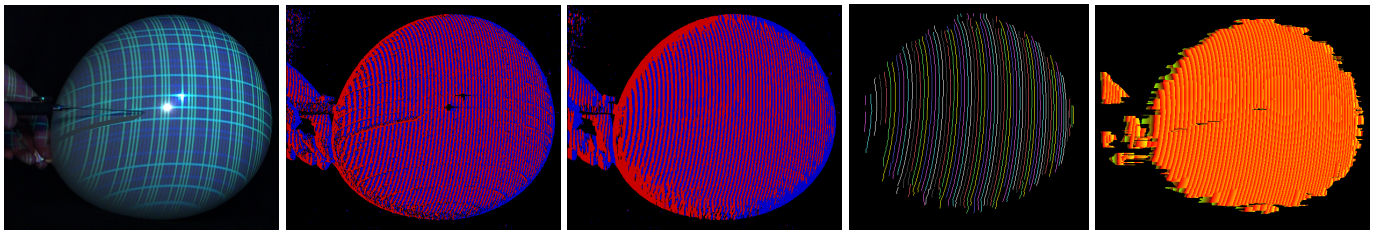
$$W_{pq}(f_p, f_q) = \min(|(f_p + d(p,q)) \bmod 8 - f_q|, \\ |8 - (f_p + d(p,q)) \bmod 8 - f_q|)^2 \tag{17}$$

where $d(p,q)$ is 1 if q is on the next horizontal curve in the $+y$ direction, -1 if q is on the next horizontal curve in the $-y$ direction, and 0 otherwise. The definition means that the IDs increase or decrease along the horizontal direction while they should be same along the other directions. The example of the result of curve detection is shown in Fig. 7(d).

C. Matching grid points between adjacent frames

Since the relative position between the camera and the projector is fixed, a grid point exists on the epipolar line even if the target object moves during the observation in the case of one projector. The corresponding points between frames, therefore, can be found along the epipolar line. In Fig.9, the candidates of the corresponding point of a grid point $\boldsymbol{p}$ at frame T are $\boldsymbol{p}'$ and $\boldsymbol{p}''$ at frame $T-1$, which are on the epipolar line of $\boldsymbol{p}$. Because each grid point has two curve IDs of the vertical and horizontal curves, the corresponding point is chosen from the one that has the same curve ID among the candidates. If multiple candidates remain, the closest one from the grid point is chosen.

In some cases, wrong correspondences can occur due to the error of curve detection or inaccurate calibration. Since the points corresponding to the grid point on a curve should be on the same curve, we remove wrong correspondences by voting the curves that the corresponding points belong to. If



(a) (b) (c) (d) (e)

Fig. 7. Curve detection based on BP. (a) is the input image. (b) and (c) show the labels by message passing after 0 and 10 iterations, respectively. Each pixel classified to the labels P (red), N (blue), and 0 (black). (d) is the result of curve detection. (e) is the phase computed by using Gabor function.

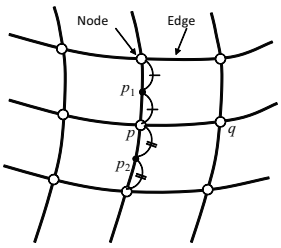


Fig. 8. The hue of a curve that consists a grid point is determined by the average of two mid points.

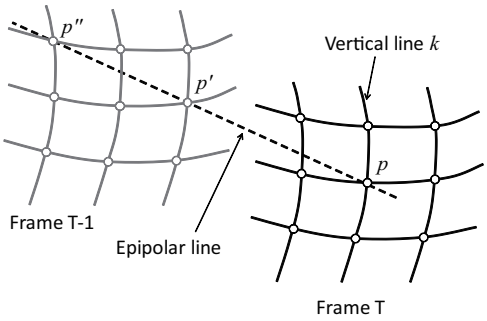


Fig. 9. The candidates of the corresponding point of a grid point $\boldsymbol{p}$ are $\boldsymbol{p}'$ and $\boldsymbol{p}''$ that are on the epipolar line of $\boldsymbol{p}$. The corresponding point is chosen from the one that has the same curve ID among the candidates.

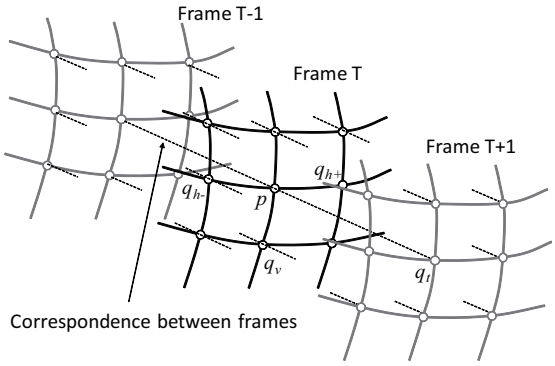


Fig. 10. If the correspondence of grid points between frames is given, the graph becomes three dimensional one. The $d(p,q)$ in Eq. 17 is 1 for q_{h+}, -1 for q_{h-}, and 0 for q_v and q_t.

the number of votes for the curve is small, the correspondence is regarded as wrong and removed. Since the correspondence for all grid points are not necessary for the proposed method, it is sufficient if grid points are connected to the adjacent frames as a graph representation to use the information for both recomputing the curve IDs and 3D reconstruction.

Once the correspondences of grid points between frames are given, the information can be used to re-compute the curve IDs by using the three-dimensional graph as shown in Fig.10. A grid point p is connected to adjacent points by temporal correspondence between adjacent frames in addition to horizontal and vertical curves. The curve IDs are computed based on BP again with three-dimensional regularization. In this case, $d(p,q)$ in Eq.(17) is 1 for q_{h+}, -1 for q_{h-}, and 0 for q_v and q_t.

In the cases of multiple projectors, the grid point is on a line different from the epipolar line, which is the projection of intersecting line of two light planes onto the camera image. Since the directions of the lines depend on the combination of pattern planes, it cannot be computed *a priori*. We therefore search the neighborhood to find the correspondence instead of searching along the epipolar line.

D. Calculating dense phase of periodic code for all pixels

Since the number of lines in a cycle is $q^n = 8$ where $q = 2$ and $n = 3$, the IDs of curves are assigned from 0 to 7. They are integer values and assigned for the pixels on the curves. The values for the pixels that are not on the curves can be computed by interpolating the curves, and it is considered as the phase of the periodic function.

Fig.11(a) is an example of the cost functions $C_p(P)$ and $C_p(N)$ of vertical curves which are computed in Sec.IV-A. The cost functions become periodic along the horizontal axis and the frequency varies according to the object's shape. The grid lines correspond to the positions where $C_p(P) = C_p(N)$ and $\frac{dC_p(P)}{dx} < 0$. By computing the offset relative to the grid lines for each pixel, the corresponding coordinate in the projector image can be obtained.

In the proposed method, the cost function is approximated by a sine curve, and the phase is used as the offset relative to a curve. Since the interval of lines is constant, the phase of each pixel for the curves is computed. The algorithm for computing phases with vertical curves becomes as follows:

1) For the pixels of which the vertical coordinate is y, find the positions of curves and compute the gap between adjacent two curves. Let $L(x)$ the gap at a pixel (x, y) as shown in Fig.11(b). If the difference of two vertical curve IDs is not 1, omit computing the phase.
2) Calculate the correlation z at the pixel (x, y) between the difference of the cost functions $C(f_n) - C(f_p)$ and a 1D complex Gabor function $g_{(x,y)}(x - x_0)$ of the wavelength $L(x)$, where the line position on the left side is x_0.
3) If $|z| > \kappa$ at the pixel (x, y), the phase $\theta(x, y)$ is given by $\theta(x, y) = \arctan(\Re z/\Im z)$. The phase ϕ that corresponds to the pixel (x, y) in the camera image is computed by $\phi = l(k + \theta(x, y)/2\pi)$, where k is the curve ID of the reference vertical line, and l is the gap

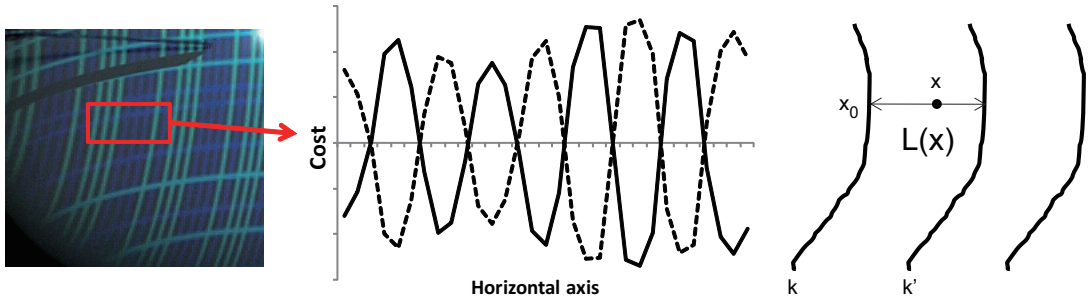


Fig. 11. (a) An example of the cost functions $C_p(P)$ (solid line) and $C_p(N)$ (dotted line) computed in Sec.IV-A for vertical lines. They become periodic functions that depend on the shape. (b) The gap $L(x)$ between detected curves is used for computing phase if the difference of two vertical curve IDs, k and k', is 1.

of vertical curves in the coordinate of projector image.

The complex Gabor function $g_{(x,y)}(t)$ is defined as follows:

$$g_{(x,y)}(t) = \exp\left(\frac{-t^2}{2\sigma^2}\right)\exp\left(\frac{-2\pi it}{L(x)}\right), \qquad (18)$$

where $\sigma = \frac{2}{3}L(x)$ and the magnitude is normalized so that $\sum_t g(t) = 1$ after discretization. If the variation of the cost function is small, which means that the estimated phase is not reliable, the proposed method computes the phase only for the absolute value of z is larger than the threshold κ. At the occluding boundary, the shape should not be interpolated. The occlusion can be detected by checking if the difference of curve IDs is 1. For the points that are not between two curves, the phase is extrapolated by using the closest curve and $L(x)$.

By considering the connection of curves, the periodic curve ID can be unwrapped in each connected component. Since the phase of each pixel can be also unwrapped, it becomes the corresponding coordinate of the projector of the pixel with a common unknown shift. Once the unknown shift is estimated by the method described in Sec.III-C, the correspondence between a camera pixel and a projector line is determined. Then, the 3D points for all pixels are calculated by triangulation.

Fig. 7(e) is the result of computing the phase from the cost function. The color from red to yellow corresponds to the phase from 0 to 2π. The green pixels represent negative phases to extrapolate the left side of curves.

V. Experiments

A. Comparison between color coding methods

To confirm the advantages of our method, we reconstructed an object with texture. In the proposed method, de Bruijn sequence of two colors (blue and green) with the code length of three was used in vertical and horizontal lines. For comparison, color coding method using only vertical color stripes [36] was also applied to the objects, in which the color pattern was constructed based on de Bruijn sequence.

The result is shown in Fig. 12. In the color coding method, 3D reconstruction was successfully achieved by using both the geometrical information obtained from the grid points and the coded IDs, even if ambiguity remains in the IDs for deciding unique correspondences. Moreover, in the proposed method, extraction of edges was less affected by the textures because of using only two colors. Standard deviation of the proposed method and the color coding method was 2.09mm and 4.15mm, respectively. Note that for calculating the standard deviation, manual process to remove many outliers was required for the color coding method, whereas nothing was required for our method. From the results, we can confirm that the proposed method provides more dense and stable reconstruction than the previous color coding method.

Fig.13 is another example of the 3D reconstruction of a textured object. Even though the colors of curves in the camera image are affected by the color checker pattern, the relative difference of the two colors can be distinguishable and the proposed method succeeded to detect the periodic curve IDs as shown in (c).

B. Removing wrong connection by using color codes

We estimated the effect of removing wrong connections by using color codes. The results of 3D reconstruction with/without removing wrong connections are compared for a scene shown in Fig.7. Since the image is captured at the moment that a balloon is bursting, the detected vertical curves are wrongly connected near the boundary between the front and back surface. Fig.14(a) and (b) show the result without removing wrong connections. In this case, the front and back surfaces of a balloon are connected. With our technique, such wrong connections are successfully cut and the shapes of the front and back surfaces are separately reconstructed as shown in Fig.14(c) and (d).

C. Reconstruction with multiple projectors

Next, we tested the 3D reconstruction with multiple projectors. In the experiments, we first used CG input images generated with a ray tracing software. We used the model of bunny from Stanford University [37] as the 3D object.

Three projectors were located at the right, left and upper sides of a camera. Fig.15(a),(b) and (c) show the input images with one, two and three projectors, respectively. The distances between the camera and projectors were 1.0, and the distance from the camera to the object was about 3.2. The field of view of the camera was about 40 degrees. The one of the projectors were shared for all the cases, which is used for the comparison of correspondence.

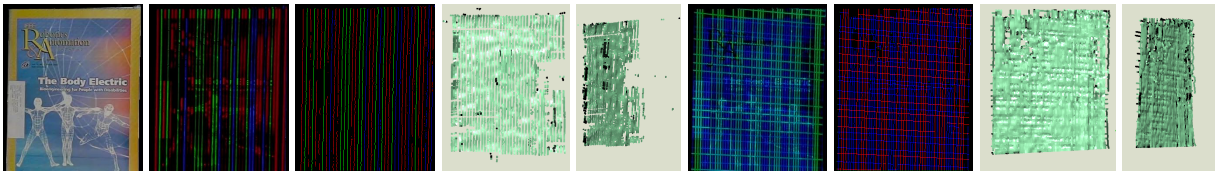


Fig. 12. Comparison between color coding method [36] and our method: (a) target object, (b) captured scene with color coding method, (c) detect curves (d)(e) reconstructed shape, (f) captured scene with our method, (g) detected curves, and (h)(i) reconstructed shape.

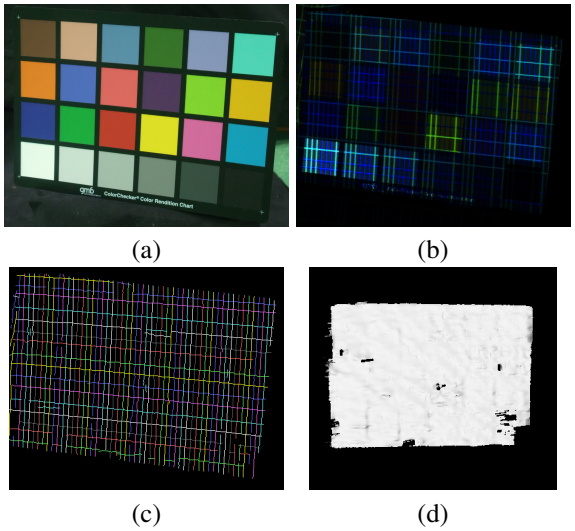


Fig. 13. The curve IDs are determined as shown in (c) by the proposed method from the captured image (b) of the color checker pattern shown in (a). The dense shape (d) is reconstructed by calculating the phase between curves.

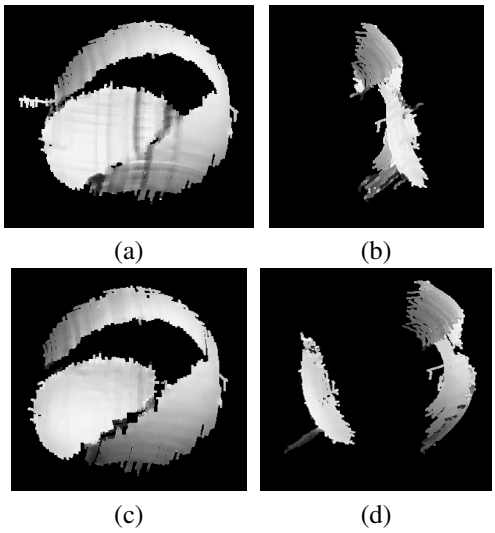


Fig. 14. (a) and (b) show the result without removing wrong connections. In this case, the front and back surfaces of a balloon are regarded as a connected surface. (c) and (d) show the result reconstructed after cutting curve between the grid points of different curve IDs. (a) and (c) are the front views while (b) and (d) are the side views.

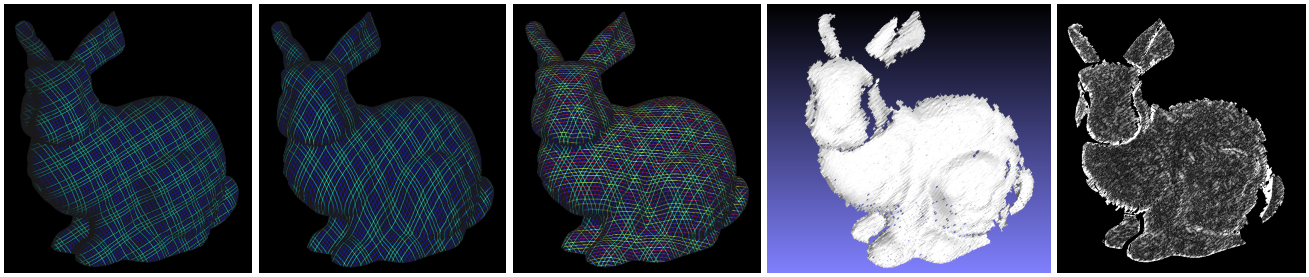


Fig. 15. The object is illuminated by (a) one, (b) two and (c) three projectors, respectively. In (a), a grid pattern is illuminated from the projector, whereas in (b) and (c), each of the projectors illuminates a parallel-line pattern. The dense shape is reconstructed as shown in (d). (e) shows the difference between the computed coordinate and the ground truth. The bright pixels indicate large errors.

We first evaluated the accuracy of correspondence between the camera and the projector obtained by calculating dense phase. The estimated correspondence for each pixel was compared with the ground truth. Fig.15(e) shows the difference between the computed coordinate and the ground truth. The bright pixels indicate large errors. The pixels at the occluding boundary had large errors, too. While the root-mean-square (RMS) error of the projector coordinate was 1.02 pixels if all pixels were used, it became 0.175 pixels if the boundary pixels were omitted from the computation. In the latter part of this section, the boundary pixels are omitted from calculating errors. This result shows that the proposed method can find the correspondence in subpixel accuracy by interpolating the phase if the surface normal is directed toward the camera and the projector enough to calculate the color code.

The error of correspondence is caused by two reasons. The first one is the error of calibration. We tested the proposed method with the intrinsic and extrinsic parameters that were estimated under imperfect correspondence; the calibration of the system was simulated by adding random noises to the point correspondence that was used for calibration. The correspondences and depths were calculated by changing the standard deviation of the noise. Fig.16 shows the RMS errors of correspondence and depth compared with the ground truth,

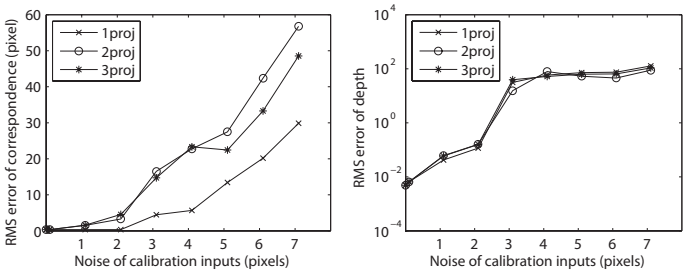


Fig. 16. The proposed method is tested with the intrinsic and extrinsic parameters that are estimated by using noisy correspondence. The calibration of the system is simulated by adding random noises to the point correspondence that is used for calibration. The errors are calculated by changing the standard deviation of noise: (a) the correspondences between camera and projector and (b) depths from the camera.

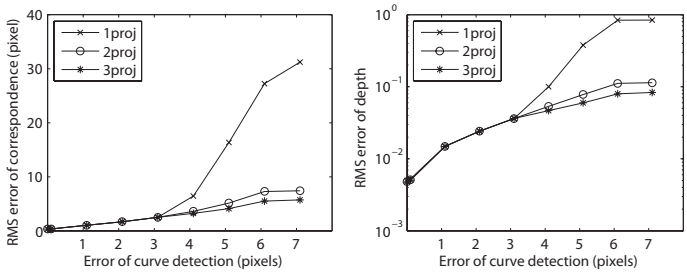


Fig. 17. The proposed method is tested by adding errors to the calculated phase, which simulates the error of curve detection and color estimation. The errors are calculated by changing the standard deviation of noise: (a) the correspondences between camera and projector and (b) depths from the camera.

which are the averages of 100 trials for each standard deviation of the noise. The errors of the results are nearly linear if the standard deviation of the noise is less than two pixels. The three cases do not have large difference with respect to the depth error as shown in Fig.16(b), while the case of one projector shows the best result with respect to the correspondence error as shown in Fig.16(a).

The second reason is the error of curve detection and color estimation. The calculated phase can shift because of the wrong detection. We tested the proposed method by adding errors to the calculated phase. The correspondences and depths are calculated by changing the standard deviation of the errors. Fig.17 shows the RMS errors of correspondence and depth compared with the ground truth, which are the averages of 100 trials for each standard deviation of error. If the error of curve detection is larger than three pixels, the error of the case of one projector increases drastically because the offsets of cycles are wrongly estimated. The errors of the cases of two and three projectors are nearly linear even with large error of curve detection. The difference between the cases of one projector and the others is that the axes of pattern planes are placed at skew position in the case of two projectors, while they intersect each other in the case of one projector. As discussed in [38], the robustness of linear solution is improved if the axes are at skew position, which is the reason why the case of two projectors gave better results. The result of three projectors shows slightly better than the case of two projectors. The reason is considered that redundant information is given to solve Eq.(7), (8), and (9) in this case.

Another advantage of using multiple projectors is to reduce occlusion. The proposed method with multiple projectors basically needs two or more projectors to illuminate a part of surface for reconstruction. Even where it is illuminated only by a projector, it can be reconstructed if it is connected to the part illuminated by two or more projectors. Fig.18 shows the result of capturing two balls by using two projectors. The vertical lines are projected from the left side of the camera, while horizontal lines are projected from the right side. Fig.18(b) is the reconstructed result. The white area can be reconstructed by the patterns of both projectors, while the red and green areas can be recontructed only by the pattern

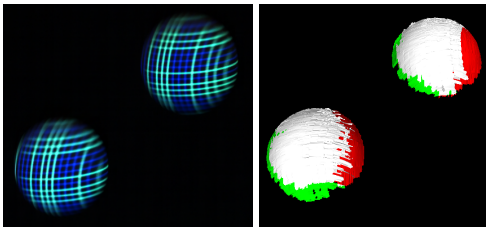


Fig. 18. (a) an input image illuminated by two projectors. (b) the reconstructed result. The white area can be reconstructed by the patterns of both projectors, while the red and green areas can be recontructed only by the pattern of one of the projectors.

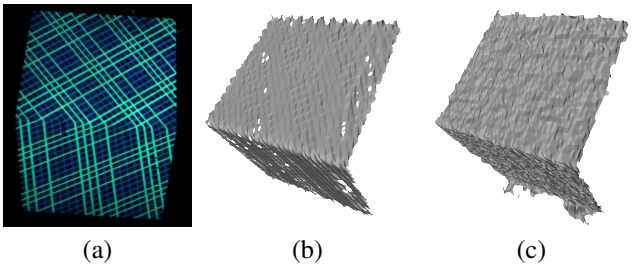


Fig. 19. A cube-shaped object is measured by the proposed method and Kinect. The size of the cube was 0.2m square and the distance from the camera was about 1.0m. (a) and (b) are the input image and the reconstructed result by the proposed method. (c) is the result of Kinect.

of one of projectors. The result with two projectors is 11% larger than that with the right projector, and 20% larger than that with the left projector.

D. Comparison of accuracy

Next, the accuracy of the proposed method was evaluated by capturing a cube-shaped object as shown in Fig.19(a). The size of the cube was 0.2m square and the distance from the camera was about 1.0m. The distance between the camera and the projector was about 0.36m. Each face of the reconstructed cube was fit to a plane to calculate RMSE. The reconstructed result by the proposed method is Fig.19(b). The average of RMSE of two planes was 0.635mm. Fig.19(c) is the result obtained by Microsoft Kinect, of which the RMSE of fitting

a plane to each face is 1.657mm. The error of the proposed method was smaller than Kinect. One of the reasons is that the baseline of the proposed method is wider than Kinect, which is approximately 0.075m. Since the proposed method is a line-based method, it is an advantage that it can find correspondences robustly even if a large image distortion occurs by a wide baseline.

E. Batch reconstruction of multiple frames

Next, we tested batch reconstruction of multiple frames with temporal connectivity given by matching grid points between adjacent frames. It is effective if spatial connectivity in a single frame is not sufficient to remove the ambiguity of correspondence. Fig.20 shows such a situation. The figures of the top row are four frames in the image sequence where many balls are falling down. The bottom row shows the results of 3D reconstruction without using temporal connectivity between adjacent frames. Since the balls are small in each frame, some of the balls in red circles are reconstructed at wrong positions due to the wrong estimation of cycles. In this experiment, six out of 76 frames have wrong estimation in the case of frame-by-frame reconstruction. In the middle row, all the frames are correctly reconstructed by batch processing of 76 frames.

F. 3D reconstruction of high frame-rate video

Finally, we conducted a dense 3D shape reconstruction using an object which changes its shape at the moment to prove the ability of high frame-rate scanning. As mentioned in the introduction, one of the advantages of one-shot 3D scanning is suitable to capture the shape of objects in extremely fast motion.

The first example is a bursting balloon with 1000 frames/second (FPS) and shutter speed 1/20000 second. Fig.21 shows the moment of the bursting balloon. While it bursted so quickly and only a several frames can be captured during the burst, the proposed method successfully generated the 3D shape for each frame. The size of an input image is 512×512 pixels, and the average time to generate the shape for a frame is 1.33 seconds. The computational cost is mainly depends on the number of grid lines projected onto the object. The projector was a 3-LCD XGA projector and an Intel Core 2 Duo 3.16GHz processor was used to compute the shapes.

Next, the shape of water splash, deforming cloth, and deforming face are captured with our method. The image sequences were captured at 60-2000 FPS by using a high speed camera. The objects are illuminated by a single projector. The image sizes of the camera and the projector are 1024×1024 and 1024×768 pixels, respectively. Fig.22 shows the four frames of the input images and results from the three sequences. In the case of the water splash (a), the water was white and opaque and the pattern was reflected on its surface. The proposed method succeeded to capture the shape of the water splash and waves caused by the ball. In the case of the deforming cloth (b), the detailed shape, such as waves and creases caused by a hit of a ball, was captured. In the case of the face (c), the cheek was hit by a hand, and we can observe the deformation from the captured shape. The average computational time was 5.00, 6.15, and 3.15 seconds for each frame by Intel Xeon 2.4GHz processor.

VI. Conclusion

In this paper, a one-shot active stereo techniques that can reconstruct dense 3D shapes are proposed. The proposed techniques realized 3D reconstruction by using dense grid patterns. For a solution, we re-formulated the problem of 3D reconstruction from a grid pattern and introduced two approaches, such as coplanarity based method and epipolar based method; both efficiently reconstruct the shape from grid pattern. We also introduced spatio-temporal constraint, by which the batch reconstruction of multiple frames is realized. Since the batch reconstruction reduces the number of parameters to be estimated, the shapes are reconstructed robustly even if the spatial connectivity is insufficient. We also proposed an efficient method to realize stable detection of dense curves of the grid pattern by classifying pixels based on belief propagation. Further, de Bruijn sequence using just two colors was proposed. The line IDs labeled by de Bruijn sequence are not only effective to remove the wrong connection of grid lines, but also used to interpolate all the pixels between lines for dense reconstruction. In the experiments, we showed the effectiveness of detecting grid pattern and the accuracy and robustness of reconstruction. Finally, the dense reconstruction of fast moving objects in a high frame-rate video was presented. In future work, we plan to utilize the reconstructed shapes of high frame-rate video for analyzing the deformable motion of objects.

Acknowledgment

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), Funding Program for Next Generation World-Leading Researchers No. LR030 (Cabinet Office, Government Of Japan), and Adaptable and Seamless Technology Transfer Program through Target-driven R&D No.AS2421093H (JST) in Japan.

References

[1] L. Zhang, N. Snavely, B. Curless, and S. M. Seitz, "Spacetime faces: High-resolution capture for modeling and animation," in ACM Annual Conference on Computer Graphics, August 2004, pp. 548–558.

[2] O. Hall-Holt and S. Rusinkiewicz, "Stripe boundary codes for real-time structured-light range scanning of moving objects," in ICCV, vol. 2, 2001, pp. 359–366.

[3] 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), vol. 27, no. 2, pp. 296–302, Feb. 2005.

[4] R. Furukawa, H. Kawasaki, R. Sagawa, and Y. Yagi, "Shape from grid pattern based on coplanarity constraints for one-shot scanning," IPSJ Transaction on Computer Vision and Applications, vol. 1, pp. 139–157, 2009.

[5] H. Kawasaki, R. Furukawa, R. Sagawa, Y. Ohta, K. Sakashita, R. Zushi, Y. Yagi, and N. Asada, "Linear solution for oneshot active 3D reconstruction using multiple projectors," in Proc. Fifth International Symposium on 3D Data Processing, Visualization and Transmission, Paris, May 2010.

[6] J.-Y. Bouguet, M. Weber, and P. Perona, "What do planar shadows tell about scene geometry?" CVPR, vol. 01, pp. 514–520, 1999.

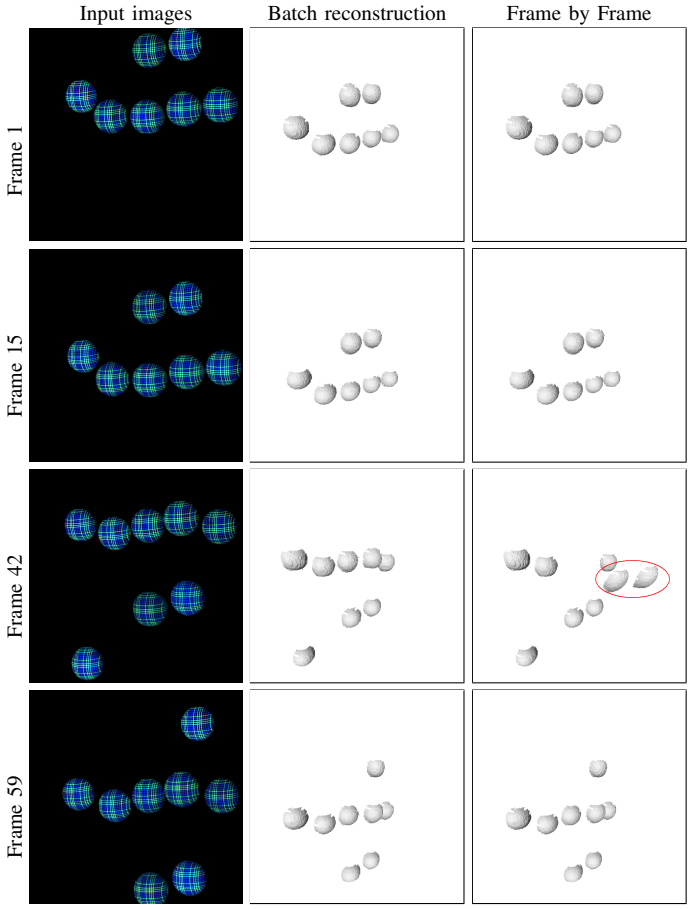


Fig. 20. The top row are the four frames in the image sequence where many balls are falling down. In the middle row, the 3D shapes of all balls are correctly reconstructed with spatio-temporal constraint between adjacent frames, while the results reconstructed frame by frame without the constraint failed to reconstruct some of the balls, which are in red circles.

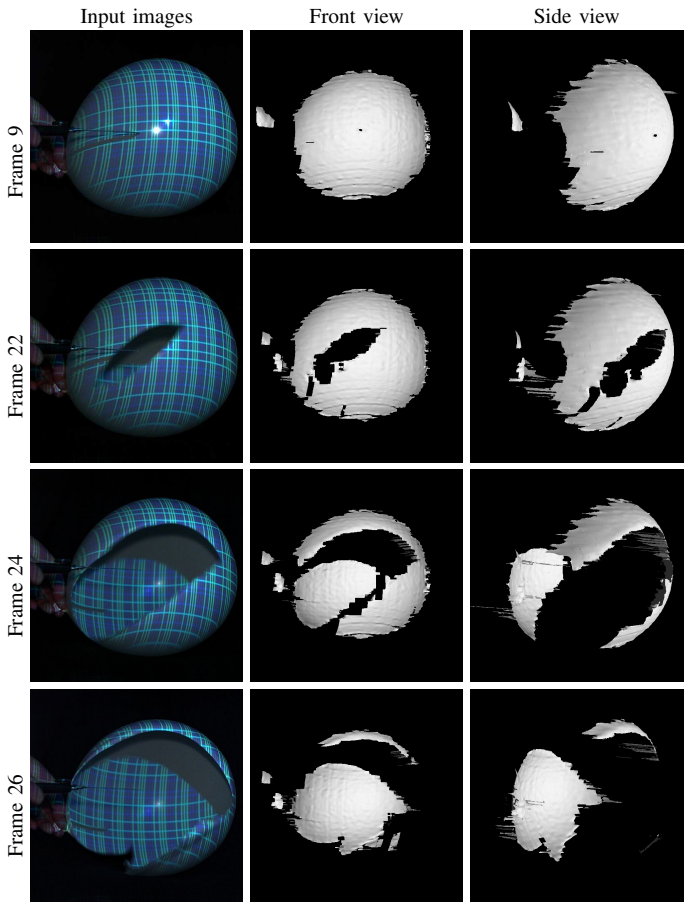


Fig. 21. 3D reconstruction of a bursting balloon. The numbers are the frame numbers in the image sequence. The size of an input image is 512×512 pixels.

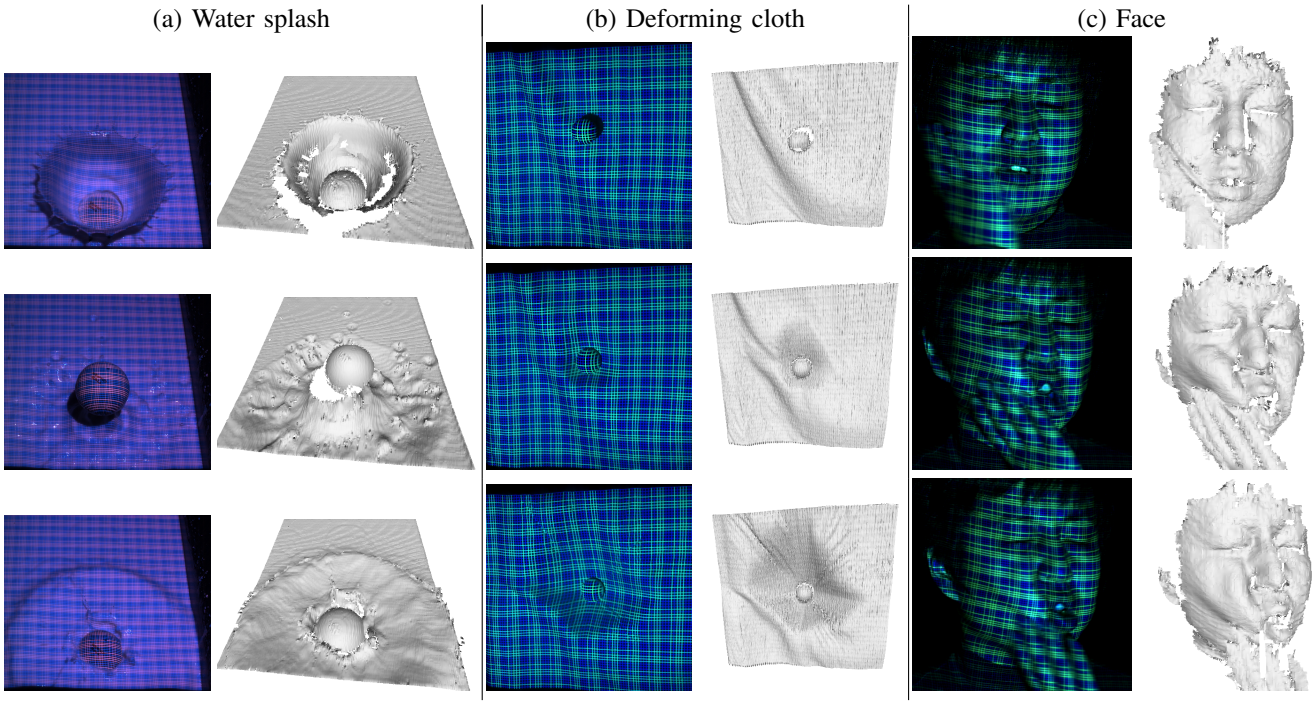


Fig. 22. The results of acquiring the shape of moving/deforming objects with a high speed camera and a projector: (a) water splash caused by a ball, (b) deforming cloth hit by a ball, and (c) deforming face hit by a hand.

[7] H. Kawasaki and R. Furukawa, "Shape reconstruction from cast shadows using coplanarities and metric constraints," in ACCVLNCS 4843, vol. II, 2007, pp. 847–857.
[8] H. Kawasaki, R. Furukawa, , R. Sagawa, and Y. Yagi, "Dynamic scene shape reconstruction using a single structured light pattern," in CVPR, June 23-28 2008, pp. 1–8.
[9] J. Batlle, E. M. Mouaddib, and J. Salvi, "Recent progress in coded structured light as a technique to solve the correspondence problem: a survey," Pattern Recognition, vol. 31, no. 7, pp. 963–982, 1998.
[10] D. Caspi, N. Kiryati, and J. Shamir, "Range imaging with adaptive color structured light," IEEE Trans. on PAMI, vol. 20, no. 5, pp. 470–480, 1998.
[11] S. Inokuchi, K. Sato, and F. Matsuda, "Range imaging system for 3-D object recognition," in ICPR, 1984, pp. 806–808.
[12] K. L. Boyer and A. C. Kak, "Color-encoded structured light for rapid active ranging," IEEE Trans. on PAMI, vol. 9, no. 1, pp. 14–28, 1987.
[13] S. Zhang and P. S. Huang, "High-resolution, real-time three-dimensional shape measurement," Optical Engineering, vol. 45, no. 12, p. 123601, 2006.
[14] T. Weise, B. Leibe, and L. V. Gool, "Fast 3D scanning with automatic motion compensation," in Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp. 1–8.
[15] S. Narasimhan, S. Koppal, and S. Yamazaki, "Temporal dithering of illumination for fast active vision," in European Conference on Computer Vision, vol. 4, October 2008, pp. 830–844.
[16] M. Young, E. Beeson, J. Davis, S. Rusinkiewicz, and R. Ramamoorthi, "Viewpoint-coded structured light," in CVPR, June 2007.
[17] C. Je, S. W. Lee, and R.-H. Park, "High-contrast color-stripe pattern for rapid structured-light range imaging," in ECCV, vol. 1, 2004, pp. 95–107.
[18] J. Tajima and M. Iwakawa, "3-D data acquisition by rainbow range finder," in ICPR, 1990, pp. 309–313.
[19] J. Pan, P. S. Huang, and F.-P. Chiang, "Color-coded binary fringe projection technique for 3-d shape measurement," Optical Engineering, vol. 44, no. 2, pp. 23 606–23 615, 2005.
[20] J. Salvi, J. Batlle, and E. M. Mouaddib, "A robust-coded pattern projection for dynamic 3D scene measurement," Pattern Recognition, vol. 19, no. 11, pp. 1055–1065, 1998.
[21] P. Vuylsteke and A. Oosterlinck, "Range image acquisition with a single binary-encoded light pattern," IEEE Trans. on PAMI, vol. 12, no. 2, pp. 148–164, 1990.
[22] Microsoft, "Xbox 360 Kinect," 2010, http://www.xbox.com/en-US/kinect.
[23] T. P. Koninckx and L. V. Gool, "Real-time range acquisition by adaptive structured light," IEEE Trans. on PAMI, vol. 28, no. 3, pp. 432–445, March 2006.
[24] C. Frueh and A. Zakhor, "Capturing 21/2d depth and texture of time-varying scenes using structured infrared light," in Proc. the 5th International Conference on 3-D Digital Imaging and Modeling, 2005, pp. 318–325.
[25] A. Ulusoy, F. Calakli, and G. Taubin, "One-shot scanning using de bruijn spaced grids," in Proc. The 2009 IEEE International Workshop on 3-D Digital Imaging and Modeling, 2009.
[26] R. Sagawa, H. Kawasaki, R. Furukawa, and S. Kiyota, "Dense one-shot 3D reconstruction by detecting continuous regions with parallel line projection," in ICCV, 2011, pp. 1911–1918.
[27] R. Sagawa, Y. Ota, Y. Yagi, R. Furukawa, N. Asada, and H. Kawasaki, "Dense 3d reconstruction method using a single pattern for fast moving object," in ICCV, 2009.
[28] R. Furukawa, R. Sagawa, H. Kawasaki, K. Sakashita, Y. Yagi, and N. Asada, "Entire shape acquisition technique using multiple projectors and cameras with parallel pattern projection," IPSJ Transactions on Computer Vision and Applications, vol. 4, pp. 40–52, Mar. 2012.
[29] P. Teunissen, "The least-squares ambiguity decorrelation adjustment: a method for fast GPS ambiguity estimation," Journal of Geodegy, vol. 70, pp. 65–82, 1995.
[30] X.-W. Chang and T. Zhou, "MILES: MATLAB package for solving Mixed Integer LEast Squares problems," GPS Solutions, vol. 11, no. 4, pp. 289–294, 2007.
[31] J. Gühring, "Dense 3-d surface acquisition by structured light using off-the-shelf components," in Videometrics and Optical Methods for 3D Shape Measurement, vol. 4309, 2001, pp. 220–231.
[32] C. Guan, L. Hassebrook, and D. Lau, "Composite structured light pattern for three-dimensional video," Opt. Express, vol. 11, pp. 406–417, 2003.
[33] C. J. Casey, L. Hassebrook, and D. L. Lau, Structured Light Illumination Methods for Continuous Motion Hand and Face-Computer Interaction. InTech, 2008, ch. Human-Computer Interaction, New Developments, DOI:10.5772/5863.
[34] P. Felzenszwalb and D. Huttenlocher, "Efficient belief propagation for early vision," IJCV, vol. 70, pp. 41–54, 2006.
[35] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," IEEE Trans. on PAMI, vol. 26, no. 9, pp. 1124–1137, 2004.
[36] L. Zhang, B. Curless, and S. Seitz, "Rapid shape acquisition using color structured light and multi-pass dynamic programming," in 3DPVT, 2002, pp. 24–36.
[37] "The Stanford 3D Scanning Repository," http://www-graphics.stanford.edu/data/3Dscanrep/, 2012.
[38] H. Kawasaki, R. Furukawa, R. Sagawa, Y. Ohta, K. Sakashita, R. Zushi, Y. Yagi, and N. Asada, "Linear solution for oneshot active 3d reconstruction using two projectors," in 3DPVT, 2010.

Ryusuke Sagawa is a senior researcher at Service Robotics Research Group, Intelligent Systems Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Japan. He received a BE in Information Science from Kyoto University in 1998. He received a ME in Information Engineering 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).

Ryo Furukawa is an associate professor 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 an research associate at Hiroshima City University in 1995, a lecturer in 2007, and an associate professor in 2012. 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, the best paper award of PSIVT in 2009, and IPSJ Best Paper Award in 2013. He is a member of IPSJ, IEICE, and IEEE.

Hiroshi Kawasaki is a professor of Department of Information and Biomedial Engineering at Kagoshima University, Japan. He received his ME on Information Engineering in 2000 and Ph.D. degree on Information and Communication Engineering in 2003 from 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 researched at INRIA Rhone-alpes, France, as a visiting professor in 2009 and Columbia university as a visiting researcher in 2011. His current research topic is on capturing shapes and textures of moving object. He also interested in photo-realistic rendering for VR/MR and ITS systems. He published over 200 research papers including ICCV, CVPR, IJCV, EG, MVA and ITSJ in computer vision and graphics areas. He also won several awards including Songde Ma Outstanding Paper Award (best paper for ACCV) in 2007, best paper award on PSIVT in 2009, Nagao Prize (best paper on MIRU) in 2011 and IPSJ Best Paper Award in 2013. He is a member of IEEE, ACM, IPSJ and IEICE.