Dynamic scene shape reconstruction using a single structured light pattern

Hiroshi Kawasaki
Faculty of Engineering,
Saitama Univ., Japan
kawasaki@cgv.ics.saitama-u.ac.jp

Ryo Furukawa
Faculty of Information Sciences,
Hiroshima City Univ., Japan
ryo-f@hiroshima-cu.ac.jp

Ryusuke Sagawa, Yasushi Yagi
Institute of Scientific and Industrial
Research, Osaka Univ., Japan
{sagawa,yagi}@am.sanken.osaka-u.ac.jp

Abstract

3D acquisition techniques to measure dynamic scenes and deformable objects with little texture are extensively researched for applications like the motion capturing of human facial expression. To allow such measurement, several techniques using structured light have been proposed. These techniques can be largely categorized into two types. The first involves techniques to temporally encode positional information of a projector's pixels using multiple projected patterns, and the second involves techniques to spatially encode positional information into areas or color spaces. Although the former 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 this problem can be resolved, however, it often uses complex patterns or color intensities, which are weak to noise, shape distortions, or textures. Thus, it remains an open problem to achieve dense and stable 3D acquisition in real cases. In this paper, we propose a technique to achieve dense shape reconstruction that requires only a single-frame image of a grid pattern. The proposed technique also has the advantage of being robust in terms of image processing.

1. Introduction

To measure 3D shapes of dynamic scenes or objects, such as human facial expressions or body motions, speed, density and accuracy of measurement are crucial. 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 for capturing dynamic scenes.

Many structured light systems temporally encode positional information of a projector's pixel into multiple patterns. Recently, structured light systems that can capture a dynamic scene by reducing the required number of patterns and increasing pattern speed have been proposed. These systems assume that there is little motion in a scene while a sufficient number of patterns for decoding are projected. In addition, the design of high-speed synchronization system is also an issue.

On the other hand, 'one-shot' structured light techniques using only single images in which positional information of the projectors' pixels are embedded into spatial patterns of the projected images have also been studied. Although the techniques can resolve the issues of rapid motions and synchronization, they typically use patterns of complex intensities or colors to encode positional information into local areas. Because of the complex patterns, they often require assumptions of smooth surface or reflectance, and the image processing tends to be difficult. If the assumptions do not hold, the decoding process of the patterns may be easily affected and leads to unstable reconstruction.

This paper presents a single 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 for the local grid pattern itself, so the pattern can be dense as long as the lines are extractable. Generally, a shape cannot be reconstructed from such a pattern. Thus, a new technique that reconstructs the grid pattern using coplanarity constraints is presented. The technique simultaneously decodes positional information of all the grid points that are connected, using constraints on coplanarity obtained from the positions of the grid points and the connectivity between them.

The proposed technique allows efficient and robust processing because it requires only local information of connectivity between adjacent grid points. In addition, it has the advantages of the shape not necessarily needing to be globally smooth as long as the local connectivity of the grid points can be extracted and thus allows the shape to be restored even when there are abrupt changes in depth due to an occlusion or in color due to texture. Moreover, decoding of the grid pattern can be achieved by a one-dimensional search, providing high-speed shape reconstruction.

978-1-4244-2243-2/08/$25.00 ©2008 IEEE

2. Related works

Shape reconstruction techniques with a structured light system, which encode positional information of a projector into temporal or spatial changes in a projected pattern, have been largely investigated and summarized in [15, 1]. A technique using only temporal changes is easy to implement, so it has commonly been used thus far [3, 8, 2]. Since the technique uses multiple patterns necessary for decoding, it requires special attention to be applied to high-speed capturing.

Techniques using only spatial encoding of a pattern allow scanning with only a single-frame image [9, 12, 13]. They typically use complex patterns or colors for phase unwrapping process and require assumptions of smooth or continuous surface or assumptions of uniform or smooth reflectance, either locally or globally. If the assumptions do not hold, the decoding process of the patterns may be easily affected and leads to ambiguities near depth or color discontinuities.

Several research reducing the required number of patterns using both temporal and spatial changes were presented [7, 15]. Although the technique is basically limited in that the scene must be static while multiple patterns are projected, Hall-Holt *et al*. [7] proposed an enhanced method to eliminate the limitation by aligning the reconstructed shape with respect to a rigid body constraint. Davis *et al*. proposed an efficient method to reduce patterns by using multiple cameras [15].

Although it does not strictly involve a structured light system, methods of shape reconstruction to include movement by spatio-temporal stereo matching are proposed [4, 16, 17]. With these techniques, a projector is only used to provide a texture that changes over time for a pair of stereo cameras to achieve high-quality depth reconstructions. All the techniques require either spatial or temporal synchronization.

Koninckx *et al*. proposed a technique allowing dense shape reconstruction based on a single image using a simple pattern, *i.e*. a 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. To cope with these problems, they developed highly sophisticated line detection and correction algorithms.

Furukawa *et al*. and Ecker *et al*. used multiple images projected by an uncalibrated line laser to construct simultaneous linear equations and reconstructed a scene or shape by solving them [6, 5]. Such equations are derived from coplanar constraints of intersection points between temporally accumulated projected line-lasers. In this paper, a grid pattern is projected onto the target scene and simultaneous

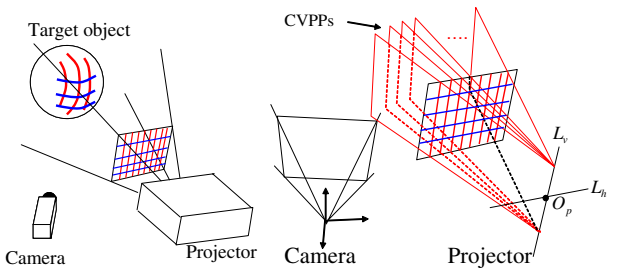

Figure 1. Scanning system:(left) the system configuration, and (right) the definition of the coordinates.

equations are constructed from the intersections of the lines on the grid using the same formulation used in those works. The projected lines are identified by solving the equations. In the proposed method, no initial identification of the grid lines are required and the lines can be dense as long as they can be extractable to acquire high-resolutional shapes. Unlike the previous 'one-shot' structured light techniques, our technique does not assume smooth shape or uniform reflectance of the surface and is robust against local discontinuities of grid patterns. Therefore, the image processing can be as simple as peak detection while vertical and horizontal scanning.

3. Shape from 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 from the projector 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 grid projected by the projector defines a plane in 3D space. Planes defined by a vertical line and a horizontal line are respectively referred to as a Vertical Pattern Plane (VPP) and a Horizontal Pattern Plane (HPP) .

The projector is assumed to have been calibrated. That is, all parameters for the VPPs and HPPs in 3D space are known. A VPP and a HPP with known parameters are referred to as a Calibrated VPP (CVPP) and a Calibrated HPP (CHPP), respectively. In addition, all CVPPs are assumed to contain the same line L_v, as in the figure 1(right). Similarly, all CHPPs contain a line L_h. The intersection of these two lines, L_v and L_h, corresponds to the optical center of

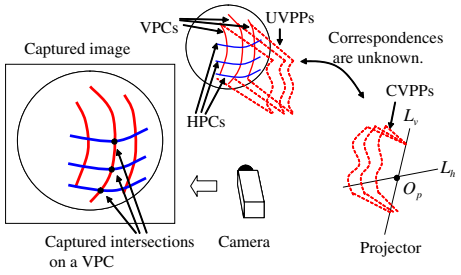

Figure 2. CVPPs and UVPPs.

the projector O_p. 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 a 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 referred to as 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. Here, these points are referred to as captured intersections. These 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 referred to as an Unknown VPP (UVPP). An Unknown HPP (UHPP) is similarly defined.

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, and the same assumption holds for UHPPs as well; this case happens when, for example, a single CVPP generates multiple VPCs because of discontinuities.

A set of intersection points that are connected by VPCs or HPCs are defined as "a linked set". A linked set can include discontinuities between adjacent grid points as long as the two points are connected by a path of multiple VPCs and HPCs as shown in Fig.3(left) and (right). By using this term, all the captured intersections are divided into multiple linked sets (*e.g.* two linked sets are observed 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

Let the CVPPs obtained by calibration be represented as $V_1, V_2, \cdots, V_M$, CHPPs be represented as $H_1, H_2, \cdots, H_N$, respectively. Also, let the UVPPs and UHPPs obtained from the captured image be represented

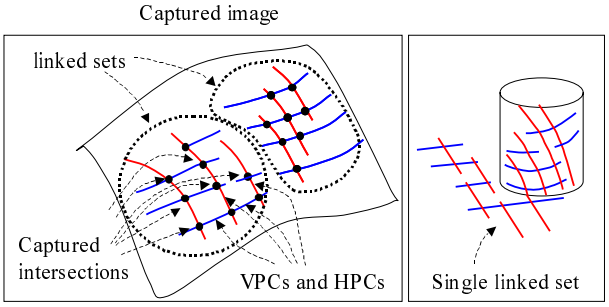

Figure 3. Linked sets of captured intersections:(left) an example of a linked set, and (right) an example of a linked set that include discontinuity of the pattern.

as $v_1, v_2, \cdots, v_m$ and $h_1, h_2, \cdots, h_n$, respectively. These symbols are used to represent correspondences between UVPPs and CVPPs. In this paper, the correspondence between the the k-th UVPP v_k and i-th CVPP V_i is represented as $v_k \rightarrow V_i$, which means v_k is identified as V_i,

The proposed method derives linear equations based on conditions of coplanarity with regard to UVPPs and UHPPs; that is, a captured intersection provides 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 UVPPs and UHPPs. These equations form a system of linear equations. In 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 (Section 3.4).

Since there is only one degree of freedom, if the position of a UVPP (or UHPP) is determined, then the positions of all the UVPPs and UHPPs will be determined. Thus by assuming a correspondence between k'-th UVPP and i'-th CVPP, which is denoted as $v_{k'} \rightarrow V_{i'}$, the positions of all UVPPs and UHPPs can be determined. Then, by comparing the positions with the CVPPs and CHPPs, each UVPP or UHPP is corresponded to the nearest CVPP or CHPP. Then, the differences between the UVPPs (or UHPP) and the corresponding CVPPs (or CHPP) are calculated. Let the sum of the squared differences be denoted as $E_{k'}(i')$, which is a function of i'. Then, the index i' that gives the minimum error function ${i'}_{min} \equiv \arg\min_{i'} E_{k'}(i')$ is searched, and the positions of UVPPs and UHPPs are determined from the correspondence $v_{k'} \rightarrow V_{{i'}_{min}}$ (Section 3.5).

3.4. Solving coplanarity constraints

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 position on the image in the coordinates of the normalized camera is $(s_{k,l}, t_{k,l})$. The planes v_k and h_l are rep-

resented by

$$a_k x + b_k y + c_k z + 1 = 0,\ d_l x + e_l y + f_l z + 1 = 0. \quad (1)$$

The location of the intersection (x, y, z) can be represented using the coordinates of the image as

$$x = \gamma\, s_{k,l},\ y = \gamma\, t_{k,l},\ z = -\gamma. \quad (2)$$

By eliminating x,y,z and γ from equations (1) and (2),

$$s_{k,l}(a_k - d_l) + t_{k,l}(b_k - e_l) - (c_k - f_l) = 0 \quad (3)$$

is obtained.

In addition, v_k contains the line L_v that contains the optical center O_p of the projector. With O_p at (P_x, P_y, P_z), and with the direction vector for L_v being (Q_x, Q_y, Q_z), we obtain

$$a_k P_x + b_k P_y + c_k P_z + 1 = 0, \quad (4)$$
$$a_k Q_x + b_k Q_y + c_k Q_z = 0. \quad (5)$$

Similarly, with the direction vector for L_h being (R_x, R_y, R_z),

$$d_l P_x + e_l P_y + f_l P_z + 1 = 0, \quad (6)$$
$$d_l R_x + e_l R_y + f_l R_z = 0. \quad (7)$$

By putting the forms (3), (4), (5), (6) and (7) together, a system of linear equations

$$\mathbf{M}\mathbf{x} = \mathbf{b} \quad (8)$$

where $\mathbf{x} = (a_1, b_1, c_1, \cdots, a_m, b_m, c_m, \cdots, d_1, e_1, f_1, \cdots, d_n, e_n, f_n)^t$ is obtained.

If all the UVPPs and UHPPs coincide with the plane containing both L_v and L_h, the aforementioned linear equation holds. This trivial solution is denoted as $\mathbf{w}$. If we decide a depth of one captured intersection, the depths of all the intersections connected to it can be determined because connected intersections are on a single plane, thus, the general solution $\mathbf{x}$ of equations (8) have 1-DOF indeterminacy. Since $\mathbf{w}$ is a particular solution and $\mathbf{x}$ has 1-DOF indeterminacy,

$$\mathbf{x} = \mathbf{w} + p\,\mathbf{u}, \quad (9)$$

where $\mathbf{M}\mathbf{u} = \mathbf{0}$. Here, $\mathbf{w}$ and $\mathbf{u}$ can be calculated by singular value decomposition (SVD) of $\mathbf{M}$ as the following. Using SVD, we can obtain the decomposition $\mathbf{M} = \mathbf{U}\ \mathrm{diag}(w_1, w_2, \cdots, w_{3n})\,\mathbf{V}^\top$, where each of $\mathbf{U}$ and $\mathbf{V}$ is a matrix with column vectors that are orthogonal with each other, and $w_1, w_2, \cdots, w_{3n}$ are singular values sorted in decreasing order. From the SVD result, $\mathbf{u}$ in form (9) can be obtained as the right-most column of $\mathbf{V}$, and a particular solution $\mathbf{w}$ in form (9) can be obtained by $\mathbf{w} = \mathbf{V}\ \mathrm{diag}(1/w_1, 1/w_2, \cdots, 1/w_{3n-1}, 0)\,\mathbf{U}^\top\mathbf{b}$.

3.5. Determining ambiguity

In the following discussion, plane parameters extracted from a solution vector $\mathbf{x}$ of the equation (8) are used. To represent the parameters, we define functions $\tilde{a_1}, \tilde{b_1}, \cdots, \tilde{f_n}$ that extracts the parameters $a_1, b_1, \cdots, f_n$ from a solution vector, respectively. For example, $\tilde{a_1}(\mathbf{x})$ is the parameter a_1 extracted from the solution vector $\mathbf{x}$.

The solution $\mathbf{x}$ includes ambiguity of the scalar p. However, by assuming a specific correspondence, for example the correspondence from the k'-th UVPP to the i'-th CVPP, we can calculate the parameter p of the form (9), and determine all the locations of UVPPs and UHPPs as the following. By assuming the correspondence $v_{k'} \to V_{i'}$,

$$A_{i'} = \tilde{a_{k'}}(\mathbf{w} + p\,\mathbf{u}) = \tilde{a_{k'}}(\mathbf{w}) + p\,\tilde{a_{k'}}(\mathbf{u}) \quad (10)$$

holds, where $(A_{i'}, B_{i'}, C_{i'})$ are the parameters of the CVPP $V_{i'}$. From this form, p can be calculated by

$$p = \frac{A_{i'} - \tilde{a_{k'}}(\mathbf{w})}{\tilde{a_{k'}}(\mathbf{u})}, \quad (11)$$

then, all UVPPs and UHPPs are determined using the p. Let this p of the form (11) be denoted as $p(k', i')$. Also, let the parameter vector (a_k, b_k, c_k) of the UVPP v_k be denoted as $\mathbf{v}_k$. Treating $\mathbf{v}_k$ as coordinates of a point, it is the "dual" of the plane v_k. Let $\mathbf{h}_l$ be defined similarly (the "dual" of the UHPP h_l). We also define $\tilde{\mathbf{v}_k}(\mathbf{x}) \equiv (\tilde{a_k}(\mathbf{x}), \tilde{b_k}(\mathbf{x}), \tilde{c_k}(\mathbf{x}))$, and $\tilde{\mathbf{h}_l}(\mathbf{x})$ similarly. Then, $\mathbf{v}_k$ and $\mathbf{h}_l$ given the correspondence $v_{k'} \to V_{i'}$, which we refer to as $\mathbf{v}_k(k', i')$ and $\mathbf{h}_l(k', i')$, respectively, can be calculated by

$$\mathbf{v}_k(k', i') = \tilde{\mathbf{v}}_k(\mathbf{w}) + p(k', i')\,\tilde{\mathbf{v}}_k(\mathbf{u}),$$
$$\mathbf{h}_l(k', i') = \tilde{\mathbf{h}}_l(\mathbf{w}) + p(k', i')\,\tilde{\mathbf{h}}_l(\mathbf{u}). \quad (12)$$

The next step is comparing the calculated UVPPs (or UHPPs) with the CVPPs (or CHPPs). For each UVPP, the difference between the UVPP and the nearest CVPP is calculated as an error. Then, we can calculate the error function of the assumption $v_{k'} \to V_{i'}$ as the sum of squared errors. By searching for the minimum of the error function, we can find the optimum correspondence and solve the ambiguity.

The comparisons are executed between UVPPs $\mathbf{v}_k(k', i')$,$(k = 1, \cdots, m)$ and CVPPs $\mathbf{V}_i$,$(i = 1, \cdots, M)$, and also between UHPPs $\mathbf{h}_l(k', i')$,$(l = 1, \cdots, n)$ and CHPPs $\mathbf{H}_j$,$(j = 1, \cdots, N)$, where $\mathbf{V}_i$ and $\mathbf{H}_j$ are parameter vectors of CVPP and CHPP that are also treated as the dual points of the planes. In this paper, comparison is done based on the squared angles between the planes. More specifically, the error function is defined as

$$E_{k'}(i') \equiv \sum_{k=1}^{m} \min_{i=1,\ldots,M} \{D(\mathbf{v}_k(k', i'), \mathbf{V}_i)\}^2 + \sum_{l=1}^{n} \min_{j=1,\ldots,N} \{D(\mathbf{h}_l(k', i'), \mathbf{H}_j)\}^2, \quad (13)$$

where D means the angle between two planes which can be defined as

$$D(\mathbf{v}_k, \mathbf{V}_i) \equiv \arccos\left(\frac{\mathbf{v}_k \cdot \mathbf{V}_i}{||\mathbf{v}_k||\,||\mathbf{V}_i||}\right). \quad (14)$$

Then,

$$i'_{min} \equiv \arg\min_{i'} E_{k'}(i') \quad (15)$$

is searched, and the set of planes $\mathbf{v}_k(k', i')$, $(k = 1, 2, ..., m)$ and $\mathbf{h}_l(k', i')$, $(l = 1, 2, ..., n)$ is the solution. The error function (13) 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 determine the ambiguity of the coplanarity constraints.

Theoretically, if there is 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 case, however, reconstruction with only an intersection may become unstable because of noise and dense grid patterns. By using multiple UVPPs and UHPPs, the endurance to noise improves because more information is used in the error function (13).

4. Configuration and detection of grid pattern

4.1. Configuration of grid pattern

As described in section 3.5, the stability of the search of i'_{min} is affected by locations of CVPPs and CHPPs. If the error function (13) has the unique minimum value at the true solution, the correct solution is obtained by the search of i'_{min}. However, if the function has multiple minimum values, the search may fail. Fortunately, the cases of multiple minimum values is very special (*e.g.* the case in which 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), and the search usually succeeds.

Another problem is that error values near minimum value usually do not have significant differences from the minimum value and it is difficult to search the true solution due to the presence of noise. A simple way to prevent this is placing CVPP and CHPP 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.

In terms of irregular patterns, to achieve more robust search, longer intervals (*i.e.* sparse pattern) are preferable, whereas dense patterns are required to densely scan an object. Therefore, it is difficult to satisfy both requirements at the same time. One effective solution is using combined patterns of dense vertical lines with uniform intervals and horizontal lines with irregular intervals (random intervals in this paper) as shown in Fig.4(left). In our implementation, vertical and horizontal patterns are respectively colored red and blue so that they were easily distinguishable.

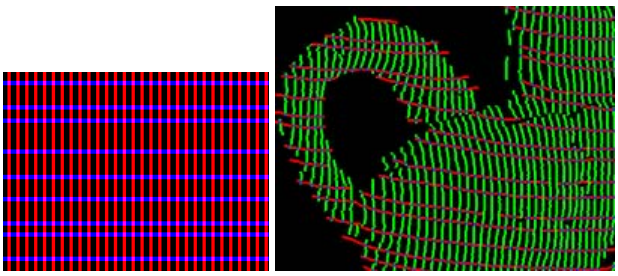

Figure 4. Example of (left) a projected pattern and (right) detected HPCs and VPCs. In the right figure, red curves are HPCs, green curves are VPCs, and intersection points are blue dots.

Note that this method does not directly encode information into the patterns, but uses their connectivity, and thus complicated encoding patterns like a Debruijn sequence [9] are not required. Just a single color and a single pixel width for lines are enough for this method. Because of those features, this method is less affected by patterns distorted by 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 intensity of the light is relatively weak compared to strong light sources like lasers or special projectors, simple thresholding techniques cannot be applied. Therefore, we implemented curve extraction algorithm based on peak detection as follows.

First, the captured image is scanned horizontally and vertically, detecting VPCs and HPCs, respectively, with subpixel accuracy using peak detection algorithm. Then, these peaks are connected using a simple labeling technique. In our experiments, above-mentioned technique sufficiently worked to retrieve satisfactory results, however, many line detection algorithms have already been proposed and can be used to obtain better results. Fig. 4(right) shows example of detected lines and intersection points.

In processing real images, several types of errors in detection of patterns may occur. One of them is disconnection of detected VPCs or HPCs. This type of error is not a serious problem because the solution of the linear equation (8) remains the same as long as the intersections configure the same liked set. Other types errors such as wrong connection of multiple VPCs (or HPCs) or wrong detection of intersections changes the solution of the linear equation. However, since the linear equation is over-constraint, our method is robust to these types of errors; the robustness on image processing is one of the most important advantages of our method.

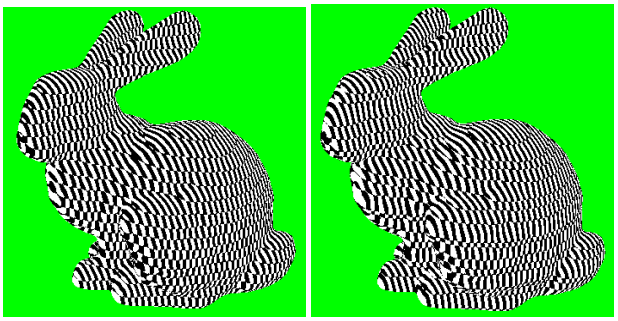

Figure 5. Synthesized grid patterns:(left) uniform intervals and (right) irregular (randomized) intervals.

5. Experiments

5.1. Simulation data

Several experiments were conducted to demonstrate the effectiveness of the proposed method.

The first experiment was conducted to confirm 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 purposefully randomized to disturb the uniformity of the grid pattern in order to increase stability of determining correspondences as described in section 4.1. The synthesized camera images of the grid patterns are shown in Fig. 5. In the images, the intervals of the VPCs were 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. In the results of both patterns, the correct correspondences for all the UHPP and UVPP were selected and the reconstructed shapes exactly matched the ground truth. The shape obtained from the data of irregular (random) intervals with the ground truth is shown in Fig. 6.

The processing time for reconstructing the input data created using the pattern with randomized intervals on a PC with a CPU of Xeon(3.8GHz) was about 1.6 sec. Almost all of the time was spent for calculation of SVD. Once $\mathbf{w}$ and $\mathbf{u}$ of form (9) are obtained from the SVD result, the minimum value search of form (15) took less than 0.001s.

Next, several experiment were 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. 5 were used and isotropic 2D Gaussian noise was added to the captured intersections. The proposed method was applied to data with various noise levels, where noise levels were standard deviations of the noise in pixels. 20 tests were conducted for each noise levels. The error ratios of the searches of i'_{min} in the form (15) are shown in Fig. 7. The results confirmed that stability of the algorithm was improved using the pattern with irregular (random) intervals.

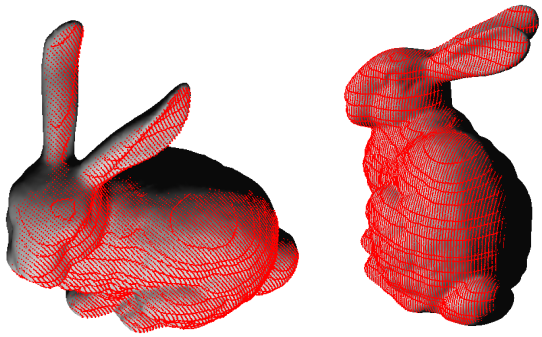

Figure 6. Result of simulation data with irregular (randomized) intervals. The red points are reconstructed points and the shaded surface is the ground truth.

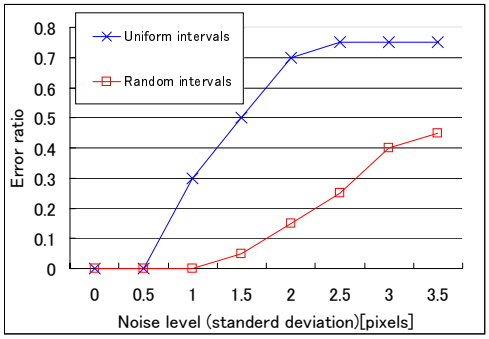

Figure 7. Error ratios with different noise levels.

Figure 8. Real 3D Scanning System

5.2. Real data - Rigid body objects

An actual 3D scanning system was built as shown in Fig. 8. Processing is performed by a PC with a Pentium Xeon 2.5Ghz CPU. Patterns were projected by a projector with a resolution of 1024×768 and scenes were captured by a CCD camera (720×480 pixels).

Figure 9 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 suc-

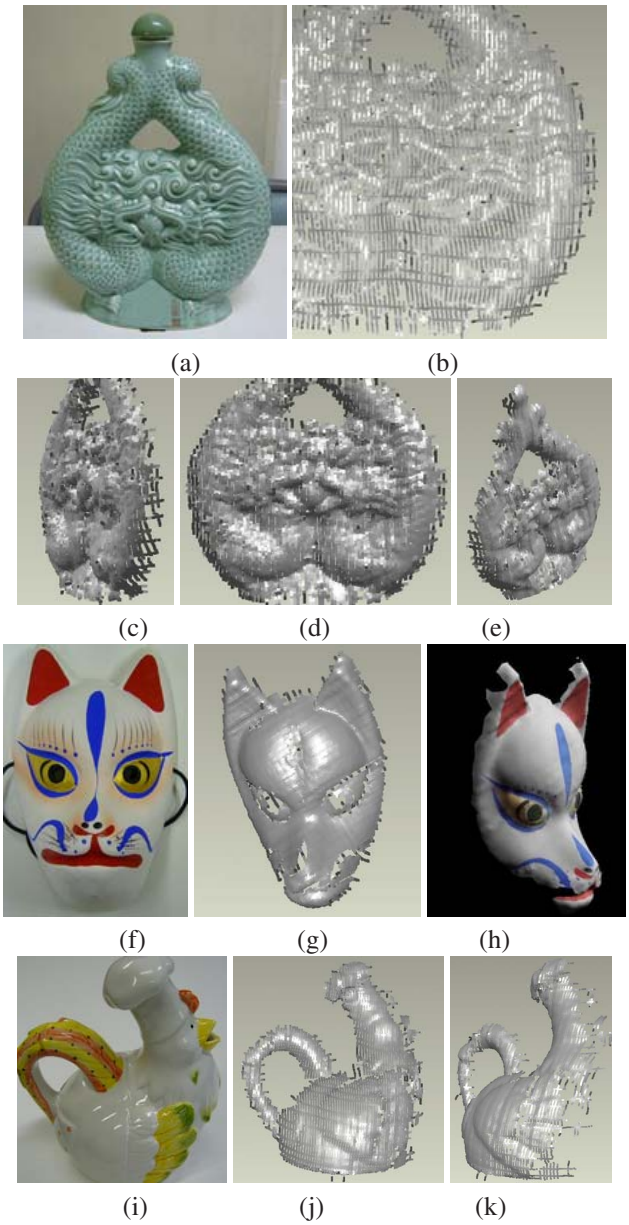

Figure 9. Reconstruction results of static objects: (a)(f)(l) the target objects, (b) close-up view of reconstructed shape of (a), (c)(d)(e)(g)(j)(k) reconstructed shape and (h) textured shape after hole-filling.

cessfully recovered with the current method. Unlike with previous methods, the proposed technique achieves reconstruction even if jump edges and abrupt color changes exist and lines are frequently segmented.

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 also reconstructed in order to evaluate the accuracy of the proposed method. The process of reconstruction was conducted in the same way as in the previous experiment, except that the 3D scene was also measured by a 3D scanner based

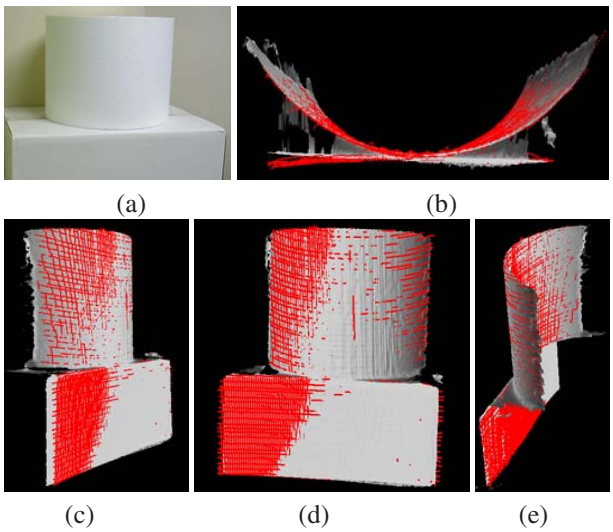

Figure 10. Reconstruction and evaluation results: (a) the captured frames and (b)-(e) the reconstructed model displayed with the ground truth data (red points: reconstructed model and shaded model: ground truth).

on coded structured light [14] as the ground truth. Figures 10(a) shows the captured scene while (b) to (e) show both the reconstruction (red points) and the ground truth (polygon mesh). Although there were small differences observed between the reconstruction and the ground truth, the shape was correctly restored. The RMS error of the reconstruction from the ground truth was $0.52mm$.

5.3. Real data - Deformable object

Finally, 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. Figures 11 show sample captured scenes and results of reconstruction. Results indicate that the proposed method successfully restored the human facial expressions with dense and accurate 3D point clouds.

6. Conclusion

This paper proposes a technique to densely measure shapes of both static and dynamic scenes or objects using a single projection of structured light. 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 local connection information of a grid pattern. Thus, it 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 (curves), vertical and horizontal, reconstruction is affected little by an object's

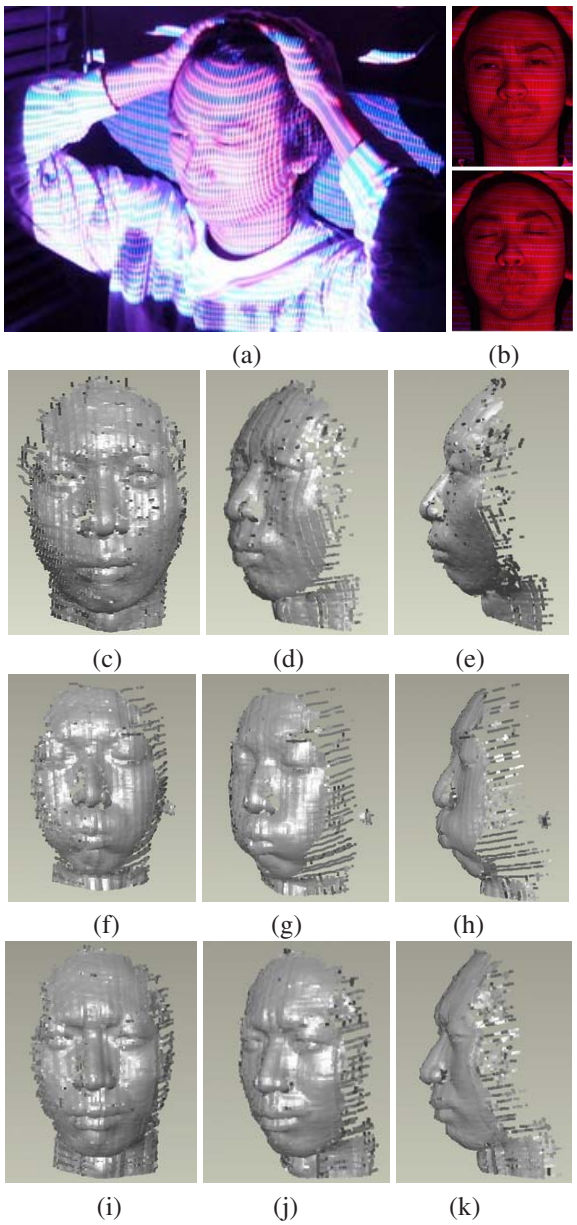

Figure 11. Reconstruction of facial expressions: (a) a capturing scene, (b) the captured frames, (c)-(e) facial expression 1(normal), (f)-(h) facial expression 2(puffing up cheeks) and (i)-(k) facial expression 3(angry). Please carefully check the differences of the shape between the eyebrows of (d) and (j).

texture, providing robust shape reconstruction. Tests were done with simulated data and real data, confirming that the proposed technique accurately reconstructed a 3D shape.

Acknowledgement

This work was supported in part by Strategic Information and Communications R&D Promotion Programme(SCOPE), 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

[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. 2

[3] D. Caspi, N. Kiryati, and J. Shamir. Range imaging with adaptive color structured light. *IEEE Trans. on PAMI*, 20(5):470–480, 1998. 2

[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. 2

[5] A. Ecker, K. N. Kutulakos, and A. D. Jepson. Shape from planar curves: A linear escape from flatland. In *CVPR*, pages 1–8, 2007. 2

[6] R. Furukawa and H. Kawasaki. Self-calibration of multiple laser planes for 3D scene reconstruction. In *3DPVT*, pages 200–207, 2006. 2

[7] 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. 2

[8] S. Inokuchi, K. Sato, and F. Matsuda. Range imaging system for 3-D object recognition. In *ICPR*, pages 806–808, 1984. 2

[9] 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. 2, 5

[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. 2

[11] T. P. Koninckx, T. Jaeggli, and L. V. Gool. Adaptive scanning for online 3d model acquisition. In *Sensor3D04*, pages 32–32, 2004. 2

[12] 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. 2

[13] 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. 2

[14] K. Sato and S. Inokuchi. Range-imaging system utilizing nematic liquid crystal mask. In *Proc. of FirstICCV*, pages 657–661, 1987. 7

[15] 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. 2

[16] 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. 2

[17] 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*, pages 548–558, August 2004. 2