IPSJ Transactions on Computer Vision and Applications Vol.4 161–169 (Oct. 2012)

[DOI: 10.2197/ipsjtcva.4.161]

Systems Paper

Capturing Textured 3D Shapes based on Infrared One-shot Grid Pattern

Kazuhiro Sakashita[1,a)] Ryusuke Sagawa[2] Ryo Furukawa[3] Hiroshi Kawasaki[4] Yasushi Yagi[1]

Received: October 31, 2011, Accepted: June 28, 2012, Released: October 19, 2012

Abstract: One of promising approach to reconstruct a 3D shape is a projector-camera system that projects structured light pattern. One of the problem of this approach is that it has difficulty to obtain texture simultaneously because the texture is interfered by the illumination by the projector. The system proposed in this paper overcomes this issue by separating the light wavelength for texture and shape. The pattern is projected by using infrared light and the texture is captured by using visible light. If the cameras for infrared and visible lights are placed at different position, it causes the misalignment between texture and shape, which degrades the quality of textured 3D model. Therefore, we developed a multi-band camera that acquires both visible and infrared lights from a single viewpoint. Moreover, to reconstruct a 3D shape using multiple wavelengths of light, namely multiple colors, an infrared pattern projector is developed to generate a multi-band grid pattern. Additionally, a simple method to calibrate the system is proposed by using the fixed grid pattern. Finally, we show the textured 3D shapes captured by the experimental system.

Keywords: shape reconstruction, textured 3D shape, line pattern, infrared light, projector-camera system

1. Introduction

Recently, several method have been proposed to capture 3D shapes of a moving object in high frame rate. If the shape of a moving object is obtained, the resulting data can be applied in various fields, such as virtual reality (VR), computer vision, gesture recognition, and robotics research. In many applications, simultaneous acquisition of texture and shape is desirable. For example, if the texture and shape of a person in motion are obtained, the model of the person can be easily rendered in a virtual world using 3D CG.

In this paper, we propose a system to capture texture and shape simultaneously based on projector-camera system. The main idea is that the proposed system acquires texture and shape by using different light wavelengths. The texture is captured by using visible light and the shape is reconstructed by using infrared (IR) light. By separating the wavelength, the system can capture the texture without affected by the projected structured light, which is a common issue for a projector-camera system.

To generate a textured 3D model from a reconstructed shape and a captured image, alignment of them are required. If the cameras for them are placed at different position, the misalignment between shape and texture inevitably occur when shape has an error even if they are precisely calibrated. To avoid the problem, we developed a multi-band camera that can capture both visible and IR lights from a single viewpoint. Another advantage of single viewpoint is that image warping using depth information is not required and rendering error near occlusion does not occur.

The method of 3D reconstruction applied in this paper is based on Segawa et al.'s method [1]. The method uses a fixed grid pattern that consists of vertical and horizontal lines and the colors of lines are determined by a periodic code. In the paper, the method used two colors from RGB (e.g., blue and green). To implement the method by using IR light, we developed a projector that projects a grid pattern by using two IR wavelengths. Additionally, since a fixed pattern mask is built in the proposed projector, it is not easy to change patterns like a video projector. Therefore, a method to calibrate the system only by using the fixed pattern is presented, which is useful to simplify the calibration step of the system.

In the following sections, related works are described in Section 2. Next, the system to reconstruct textured 3D shapes by grid pattern is briefly explained in Section 3. Then, we describe a system to capture texture of visible light and IR structured light simultaneously in Section 4. Finally, an experimental system is tested to obtain a textured 3D model in Section 5.

2. Related Work

Techniques to acquire the shape of an object are categorized roughly into to two types, active and passive methods. Stereo vision [2] and visual hulls [3] are often used in passive methods. Since these methods measure the shape of an object using multiple synchronized cameras, they are suitable for use in capturing texture and shape simultaneously. Moreover, it is possible to capture data at a high frame rate because all the input data is obtained at the camera frame rate. Although the accuracy and stability of the measured shape is low compared to active methods;

1 Osaka University, Ibaraki, Osaka 567–0047, Japan
2 National Institute of Advanced Industrial Science and Technology, Tsukuba, Ibaraki 305–8569, Japan
3 Hiroshima City University, Hiroshima 731–3166, Japan
4 Kagoshima University, Kagoshima 890–8580, Japan
a) kazuhiro-sakashita@aist.go.jp

IPSJ Transactions on Computer Vision and Applications Vol.4 161–169 (Oct. 2012)

it is sufficient to produce a 3D model for computer graphics. One drawback of a visual hull is that a large system is required since many cameras are necessary to observe an object from various viewpoints.

For active methods, laser range finders and structured-light methods are commonly used. Especially, a laser range finder is often used for an accurate measurement of a shape. Laser range finders measure the distance between the sensor and an object based on triangulation or time-of-flight. A time-of-flight sensor that changes the direction of a laser beam by controlling a mirror cannot capture shapes at a high frame rate, because each change takes several seconds or minutes. Although fast range finders [4], [5] have recently been developed, their resolution and accuracy are not sufficient due to the trade-off between time and accuracy.

Structured-light methods cast light patterns on an object using a projector and a reflected patterns are observed by a camera. This technique determines the corresponding coordinates between the projector and the structured-light patterns observed by the camera. The distance from the camera to the object is computed by triangulation based on the correspondences. Some fast methods [6], [7], [8], [9] are proposed based on structured-light methods and they can measure shape at a high frame rate. Especially [8], [9] are suitable for real-time acquisition because they can measure a shape from a single image. However, these methods have difficulty in obtaining texture simultaneously because visible light patterns are projected to measure the shape. Two solutions are proposed for the problem: time-sharing methods [10], [11], [12], [13], [14], [15], [16] and band-separating methods [17], [18], [19].

The time-sharing method uses structured light not only for capturing shape but also for illuminating objects to obtain texture. If the structured light patterns are removed from the images, they can be used as texture images. This is accomplished by computing the average of the images of time-multiplexing patterns, or low-pass filtering of fringe patterns. The advantage of this approach is that texture and shape can be obtained using the same camera. However, since structured light is used for both shape and texture, other illumination in the environment cannot be used. Consequently, the illumination in which these methods are applicable is restricted. Additionally, since the duration of shape and texture capture are different, misalignments between texture and shape may appear if motion is fast.

Zhang and Yau proposed a method based on phase shifting technique [10]. Since the phase shifting requires to capture multiple images for measuring a shape, it is dificult to capture the object in fast motion. For such case, one-shot scanning method is suitable; the method uses multiple colors instead to obtain 3D shapes and band separation method to obtain textures.

The band-separating methods uses different light wavelengths for shape and texture. The shape is obtained by using near-infrared (NIR) light, while the texture image is captured with visible light. Since the structured light patterns for NIR light are not detected by the camera for visible light, the acquisition of shape and texture can be accomplished simultaneously. Frueh and Zakhor [17] use vertical stripes and a horizontal line of IR. The vertical stripes are used for shape reconstruction and the horizontal line for identifying the vertical stripes. Hiura et al. [18] proposed a laser range finder that obtains a range image at 30 Hz by fast scanning of IR light stripes. A texture image is captured simultaneously with the system. Microsoft Kinect [19] can capture a shape by using an IR projector and a camera, and a texture by another CCD camera at the same time. Although there is a similarity between Kinect and ours such as shape reconstruction from a single image, there are two large differences exist; i.e., accuracy of our system is much higher than Kinect and a shape and texture has a single viewpoint with our system.

3. Acquisition of Textured 3D Shapes by One-shot Structured Light Method

3.1 One-shot 3D Reconstruction Based on Oneshot Grid Pattern

In this paper, the reconstruction algorithm is based on the work proposed by Sagawa et al. [1]. 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 projected pattern is fixed and does not change, so no synchronization is required. A grid pattern that consists of vertical and horizontal lines is projected from the projector and captured by the camera.

The flow of the algorithm is shown in **Fig. 2**. In the first step, the curves illuminated by the grid lines are detected from the captured image. In the second step, the intersection points of the detected curves (grid points) are extracted.

Each line emitted from the projector sweeps a plane in the 3D space. This plane is called a pattern plane in this paper. From the

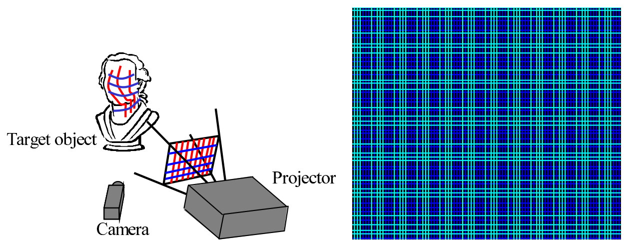


Fig. 1 (Left) Scanning system: multiple lines are projected and their intersections are detected and used for reconstruction. (Right) projected pattern.

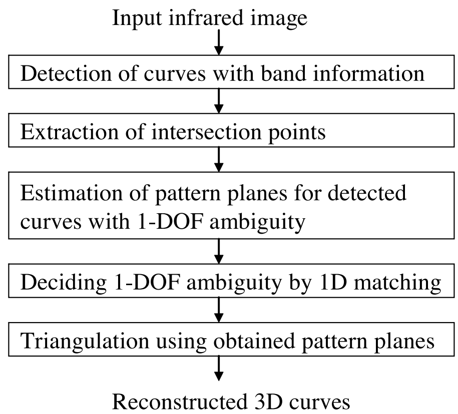


Fig. 2 Flow of the reconstruction algorithm.

extracted grid points, simple linear constraints about the crossing pattern planes can be acquired.

The following equation is obtained from the constraints:

$$\mathbf{T}\vec{\eta} = \mathbf{R}\vec{\rho} \qquad (1)$$

where **T** and **R** are matrices constructed from coordinates of grid points, $\vec{\eta}$ and $\vec{\rho}$ are vectors from the parameters of vertical and horizontal planes, respectively. Refer Ref. [20] for the detail of the formulation. In the third step, solutions of the pattern planes are obtained by solving the linear equations given by the constraints.

In the solution, there remains 1-DOF ambiguity which cannot be solved from information of the grid points [21]. In the fourth step, the ambiguity is solved from estimation of matching between the 1-DOF solution and the actual positions of the pattern planes obtained from the projected pattern and the calibration results. Intuitively, this matching process is equivalent to finding correspondences of the grid points using epipolar constraints that are consistent for the whole set of connected grid points. In the final step, the 3D information of the detected curves are calculated by triangulation using the obtained pattern planes.

In the matching process, additional information attributed to each line pattern is effective for avoiding ambiguous matching, where multiple candidates of the solution remains. In this algorithm, we use color information. To achieve this, the grid lines are colored by a periodic (thus, not globally unique) coding of de Bruijn sequence. In the reconstruction process, the periodic IDs are decoded and assigned to each detected curves, and the IDs are compared to the projected patterns at the matching stage.

Once the correspondence between a line of the projected pattern and a curve in the camera image is determined, the 3D positions of points in a curve is computed. Since it is beyond the scope of this paper to describe the detail of the algorithm, the interested readers can refer Ref. [1] for more details.

3.2 Infrared System for Capturing Textured 3D Shape

The proposed system extends the above method to capture a 3D shape with texture. The system uses an IR projector and a multi-band camera as shown in **Fig. 3**. The IR projector casts a fixed grid pattern that consists of vertical and horizontal lines. The lines have two different patterns using two NIR lights of 765 nm and 850 nm instead of two-color lights, respectively. The multi-band camera acquires images of both visible and IR wavelengths

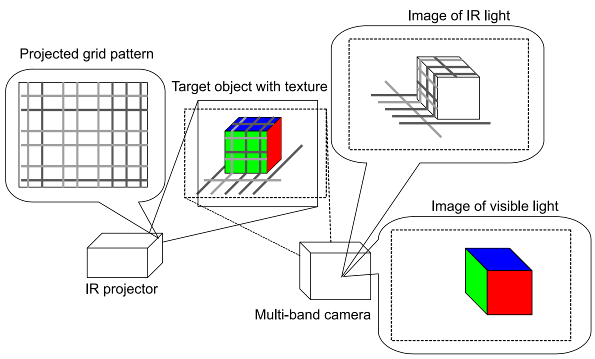


Fig. 3 The proposed system to capture textured 3D shape consists of an IR projector and a multi-band camera.

from a single viewpoint. For capturing a texture, an external light source is used for illumination. If the external light source contains an IR component, it is cut-off by a band-pass filter to avoid interference when capturing the shape. Although it is possible to increase the number of bands (e.g., by adding prisms), two bands of IR patterns are sufficient using the proposed methd. It is one of the advantages of the system because using more prisms and colors decreases light intensity and increases the complexity of both the system and the algorithm.

Since NIR lines for shape reconstruction have some intervals, the shapes between lines are linearly interpolated in our system to meke texture-mapped object by images with visible light. In addition, since the texture image is acquired from the same viewpoint, the occlusion problem does not occur in the texture mapping process in the proposed system.

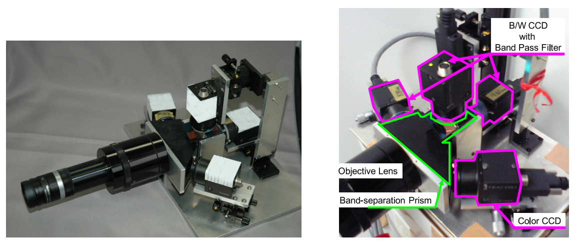
4. Proposed System for Capturing Texture and Shape

4.1 Multi-band Camera to Acquire Visible and IR Images from a Single Viewpoints

To generate a textured 3D shape, it is desirable to acquire texture and shape from the same viewpoint because the misalignment between shape and texture can occur due to the error of shape if they are capture from different positions. For solution, the proposed camera has imaging devices for visible and IR lights that are placed coaxially.

Figure 4 shows an experimental multi-band camera system. The size of the system is approximately W300 × D300 × H200 mm. The system uses three monochome CCD cameras (Point Grey FL2-03S2M-C) and a color CCD camera (Point Grey FL2-03S2C-C) with a maximum frame rate of 30 fps. The object lens consists of a general C-mount lens and a relay lens. Different fields of view can be obtained by changing the C-mount lens. The relay lens has a long back focus so that incoming ray focuses on the surface of imaging sensors after going through the optical components shown in Fig. 4 (b).

Figure 5 shows the internal structure of the multi-band camera. The optical components are composed of an objective lens, a band-separation prism, and imaging sensors. The incoming light passing through the objective lens is first split into visible, 765 nm and 850 nm IR light by the band-separation prism; a Philips prism that consists of four glass blocks, of which interfaces are coated with membranes to separate different bands of light. Although the prism can separate out 935 nm light from the other wavelength,



(a) (b)

Fig. 4 (a) an experimental system of the proposed multi-band camera. (b) the zoom up of the optical components. The object lens consists of a general C-mount lens and a relay lens. Different field of view can be obtained by changing the C-mount lens.

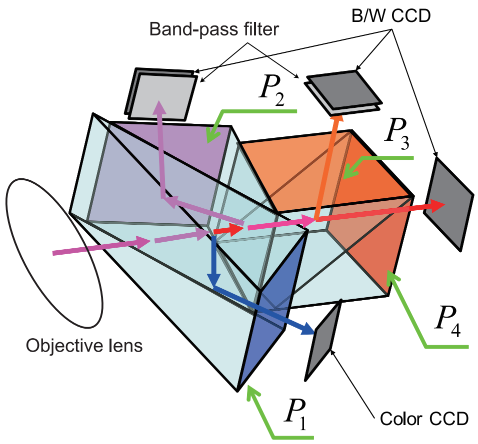


Fig. 5 The internal structure of the multi-band camera in the proposed system: The camera consists of an objective lens, a band-separation prism, and image sensors.

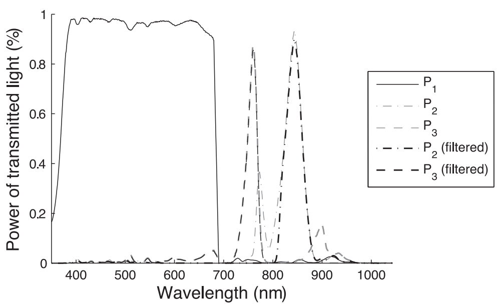


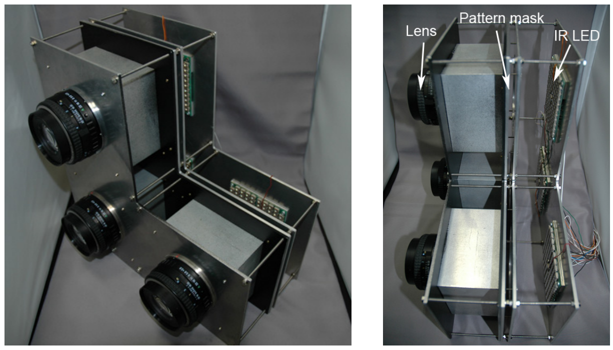
Fig. 6 More than 80% of the light power is transmitted to desirable exit faces when the IR light shown in Fig. 9 and white visible light enter from the objective lens.

935 nm light is not used in this paper because two wavelengths are sufficient for the current encoding of patterns. The wavelength will be used if the pattern is encoded by using three wavelengths. The exit faces P_1, P_2, P_3 and P_4 in Fig. 5 corresponds to visible, 850 nm, 765 nm, and 935 nm, respectively.

Figure 6 shows the power of light transmitted to each exit face when the IR light cast from the IR projector and white visible light from external light sources enter from the objective lens. P_1, P_2 and P_3 indicates the prism faces illustrated in Fig. 5. Since the results of P_2 and P_3 have crosstalk each other, band-pass filters are used to remove this. The efficiency of the light power is more than 80%. The optical path lengths from the objective lens to the four image sensors are set to be equivalent so that the rays from the same point focuses on the imaging sensors. Two monochrome cameras are used to obtain IR images as imaging sensors, and a color camera is used for visible light in this system [*1].

Akasaka et al. [22] proposed a similar system that captures IR and visible lights from the same viewpoints. The disadvantage of their system was inefficiency in terms of light energy. Since a simple beam splitter that split incoming light regardless of its wavelength was used in the system, the power of the structured light was reduced to less than half of the power discharged from the projector when splitting and combining light. Consequently, it needs strong light source to acquire image sequence at high frame rate. The proposed system has better performance in terms of the efficiency of light power by modifying the optical components of the camera, and accomplish to capture moving objects at high frame rate.

*1 One of monochrome cameras is not used in this paper, because only two IR wavelengths are used by the proposed projector.



(a) (b)

Fig. 7 The new experimental IR projector system with three arrays of infrared LEDs, pattern masks printed on transparent films, and objective lenses.

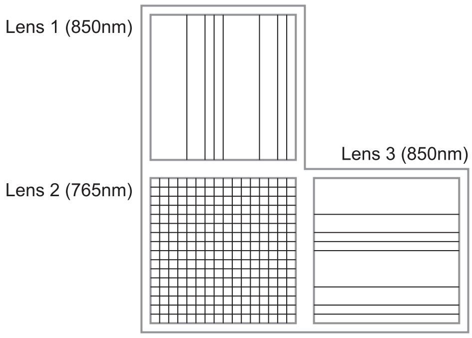


Fig. 8 An example of the projected grid pattern is formed by three line patterns that corresponds to the three lenses.

4.2 Infrared Projector to Cast Grid Pattern

To design an IR projector for 3D reconstruction, there are two requirements exist as follows: 1) the projector casts a fixed grid pattern that consists of vertical and horizontal lines. 2) two IR wavelengths are required to encode the pattern. One idea to realize them is to modify a video 3-LCD projector commercially available to cast IR lights instead of visible lights. However, since the optical components of the projectors are specialized for visible light and usually cut-off filter for IR is installed, it is not possible to modify a video projector for our purpose.

Because the pattern used in the proposed system is static, it is not necessary to change it during scanning process. Therefore, a fixed pattern mask printed on transparent film is sufficient for the system instead of LCD filters used in video projectors.

Another issue is to cast two IR light of different wavelengths. A usual LCD projectors combines lights of different colors by a combination prism before casting through the objective lens. In the proposed system, however, the pattern is simply composed of vertical and horizontal lines. Therefore, it is possible to form the pattern without combining rays by a prism. The proposed system casts lines by using three lenses placed like L-shape as shown in **Fig. 7**. The size of the projector is approximately W300 × D200 × H300 mm.

An IR LED array and pattern mask are placed behind each lens. **Figure 8** shows an example of the pattern masks for three

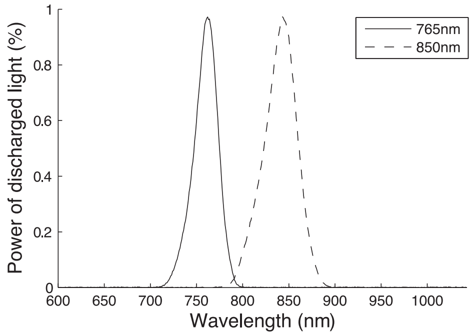


Fig. 9 The power of discharged light from each lens: the maximum of each band is normalized to one.

lenses. 765 nm light is cast through the lens 2, and 850 nm light is through the lens 1 and 3. If lines cast from the lenses 1 and 2 are parallel to the direction from the lens 1 to lens 2, the pattern planes cast from these lenses are the same. Thus, the lines are encoded by two wavelengths of IR light without combining them before emitting from the objective lenses. This approach contributes the efficiency of light power. The patterns of 850 nm are determined based on de Bruijn sequence and aligned to that of 750 nm.

Each IR LED array used as a light source has 116 LEDs and the power of each array is 0.96–1.20 W, which is much smaller than the light sources used in commercial video projectors. **Figure 9** shows the power of discharged light from each lens, of which the maximum is normalized to one. Since the bandwidth of LEDs is quite narrow, the crosstalk between different wavelengths does not occur; crosstalk is one of the main issue of image processing when a visible light source such as a commercial video projector is used.

4.3 System Calibration

4.3.1 Calibration between Image Sensors

For texture mapping, the texture image must be geometrically calibrated with the system to acquire the shape. If the texture is captured from a viewpoint different from the camera to obtain the shape, the corresponding point between the texture image and shape cannot be determined a priori. On the other hand, since the proposed system can obtain texture and shape from the same viewpoint, the corresponding points are determined by calibrating the geometry between the image sensors.

In the proposed multi-band camera, although the different wavelengths are separated by a prism, the optical structure for all image sensors is virtually the same except the positions of sensors relative to the other optical components. Since the positions of the image sensors have some small misalignments relative to the other optical components, an incoming ray are not projected on the same pixel of image sensors. The situation is illustrated in **Fig. 10**. Since the virtual optical path are common for all image sensors, the calibration to compute the corresponding points can be done by estimating the homography between the image sensors.

Estimating a homography is similar to the method used in the camera calibration. First, the image of a planar object, such as a flat plate with checker board pattern, is captured by all sensors.

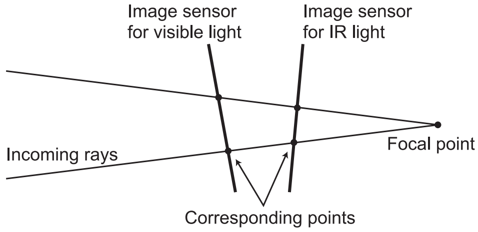


Fig. 10 The virtual optical path are common for all image sensors. Due to the misalignments, an incoming ray is projected on the different pixel of the image sensors.

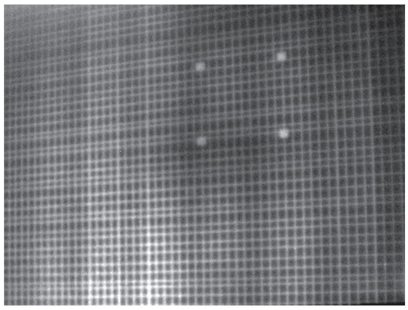


Fig. 11 The pattern of one-shot 3D reconstruction is used for calibration. The absolute coordinate of the projector is manually given by using the marker in the pattern.

By detecting four or more feature points on the object, the homography can be estimated [23].

4.3.2 Calibration of Projector-camera System

To calibrate a projector-camera system, the correspondence of the coordinates between the projector and the camera is required. It can be accomplished by capturing coded images projected from the projector. Time-multiplexed pattern, such as a pattern based on Gray code, is often used to obtain the absolute coordinates of the projector by observing the multiple coded images of the structured light. However, since the proposed system uses a fixed pattern printed on a transparent film, it is not easy to use time-multiplexed pattern by changing the pattern mask. Thus, we utilize the pattern of one-shot 3D reconstruction for calibrating the system. Since the pattern is a periodic pattern, the absolute coordinate of the projector is not automatically determined. For solution, we manually decide the first point to determine the absolute coordinate. Then, remaining points are automatically determined by using the connecting information of curves; those curves are automatically detected by efficient curve detection algorithm proposed in Ref. [1]. To ease the manual process to decide the first point, some markers are added on the pattern as shown in **Fig. 11**. We make the markers to be small so that they do not interfere the curve detection algorithm.

Once the corresponding coordinates between the camera and projector are obtained, the intrinsic of the camera and projector, and extrinsic parameters between them can be estimated as calibration of a stereo system. In this paper, an implementation based on bundle adjustment provided by Snavely [24] is used for calibration. If both calibration of the system and 3D reconstruction is accomplished only by a fixed pattern, we can simplify the projector drastically compared to commercial video projectors, which need optical devices to generate a pattern by LCD or DMD.

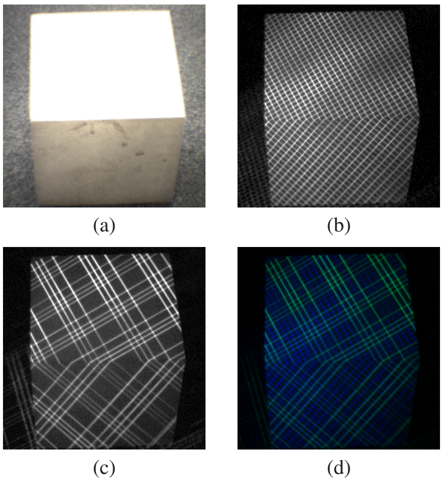


Fig. 12 Acquired images of a cube-shaped object: (a) color image of visible light (b) the projected pattern of 765 nm wavelength (c) the projected pattern of 850 nm wavelength (d) the grid pattern generated by combining 765 nm and 850 nm.

5. Experiments

We first estimate the accuracy of the experimental system by measuring a known object. The shape of a cube-shaped object shown in **Fig. 12** is acquired. Figure 12 (a), (b) and (c) are the images of visible, 765 nm IR, and 850 nm IR lights, respectively. Figure 12 (d) is the grid pattern generated by integrating the two IR images. Blue and green lines correspond to 765 nm and 850 nm lights. The images are aligned each other by using homographies estimated by the method described in Section 4.3.1. The cube is 0.2 m on a side and the distance from the camera is about 0.8 m. There is no crosstalk between visible, 765 nm and 850 nm bands thanks to the prism of band separation.

Figure 13 (a) shows the reconstructed result of a cube by the proposed method. The distance between the camera and the projector is 0.72 m and the camera and object is about 1.00 m in this condition. To evaluate the accuracy of the shape, the two faces of the cube are fit to planes, respectively. Because the reconstructed shape around the boundaries has large errors compared to the internal regions, the errors are evaluated with and without the boundaries as shown in Fig. 13 (a) and (b). The root mean square errors (RMSE) of the points from the fit planes are 0.81 mm and 0.39 mm with/without the boundary, respectively. The angle of the two planes is 88.6 degrees (the ground truth is 90.0). Figure 13 (c) shows the reconstructed result of the cube by Kinect. The RMSE of Kinect are 2.16 mm and 1.78 mm with/without the boundary, respectively. Figure 13 (d) shows the reconstructed result of the cube by time-multiplexing method (Gray code + phase shifting). The RMSE of the time-multiplexing method are 0.53 mm and 0.34 mm with/without the boundary, respectively.

The RMSE is sufficiently small compared to the distance between the camera and the object. The current implementation of calibration does not consider the lens distortion of camera and projector. The angular error can be improved by incorporating distortion parameters in future work.

Next, we show accuracy comparison of proposed method with Microsoft Kinect. **Figure 14** is the results of 3D reconstruction

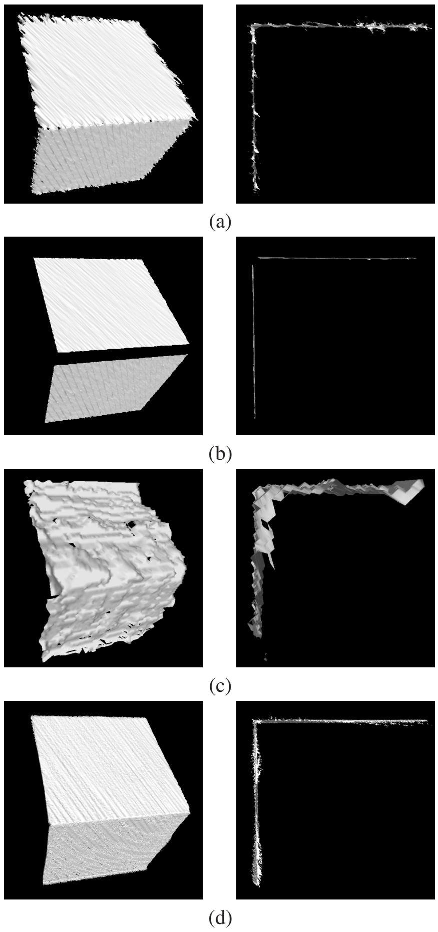


Fig. 13 The reconstructed result of the cube-shaped object. (a) is the reconstructed shape by proposed method. (b) is the shape after the boundaries of planes are removed. (c) is the reconstructed shape by using Kinect. (d) is the reconstructed shape by the time-multiplexing method.

of a mannequin. The height of the upper body is 0.40 m and the distance from the cameras is about 1.00 m. The distance between camera and projectorin the condition is 1.02 m.

Figure 14 (a) shows the shape reconstructed by a method based on time-multiplexing (Gray code + phase shifting). The shape in (b) is obtained by using Kinect. (c) shows the result of the proposed method. The result by Kinect has stepwise errors, which are due to the error of correspondence in sub-pixel accuracy. We guess it is because Kinect has a narrow baseline compared to the distance and the projected pattern is a set of points. Meanwhile, the proposed method is based on line patterns, which has an advantage to find correspondence in sub-pixel accuracy as many of range scanners is based on light sectioning. For numerical comparison, we used the result by time-multiplexing as a reference. The RMSE between (a) and (b), and that between (a) and (c) are 1.65 mm and 0.60 mm, respectively. The values are calculated by using a method proposed by Cignoni [25]. The proposed method generates a shape sufficiently close to the method of high-accuracy. As qualitative evaluation, the result by the proposed method has smooth surfaces, which shows the line positions are computed accurately.

Next, we show other examples of capturing textured 3D models. **Figure 15** is the result of 3D reconstruction with texture mapping of a head model and a stuffed bear. The figures are the im-

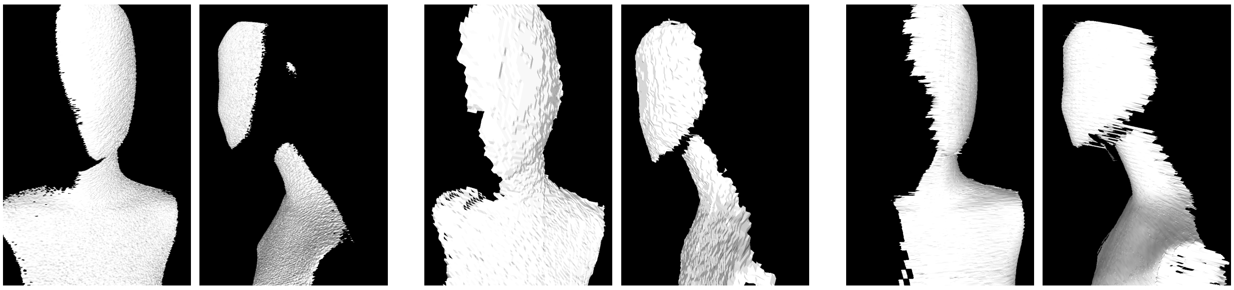


Fig. 14 The upper body of a mannequin is reconstructed for comparison. (a) is the shapes reconstructed by a time-multiplexing method (Gray code + phase shifting). (b) is obtained by using Kinect. (c) is reconstructed by the proposed method,

Fig. 15 3D reconstruction with texture mapping of a head model (left) and a stuffed bear (right). The figures are the image of visible light, the IR image, the shape without texture, and the textured 3D shape, respectively.

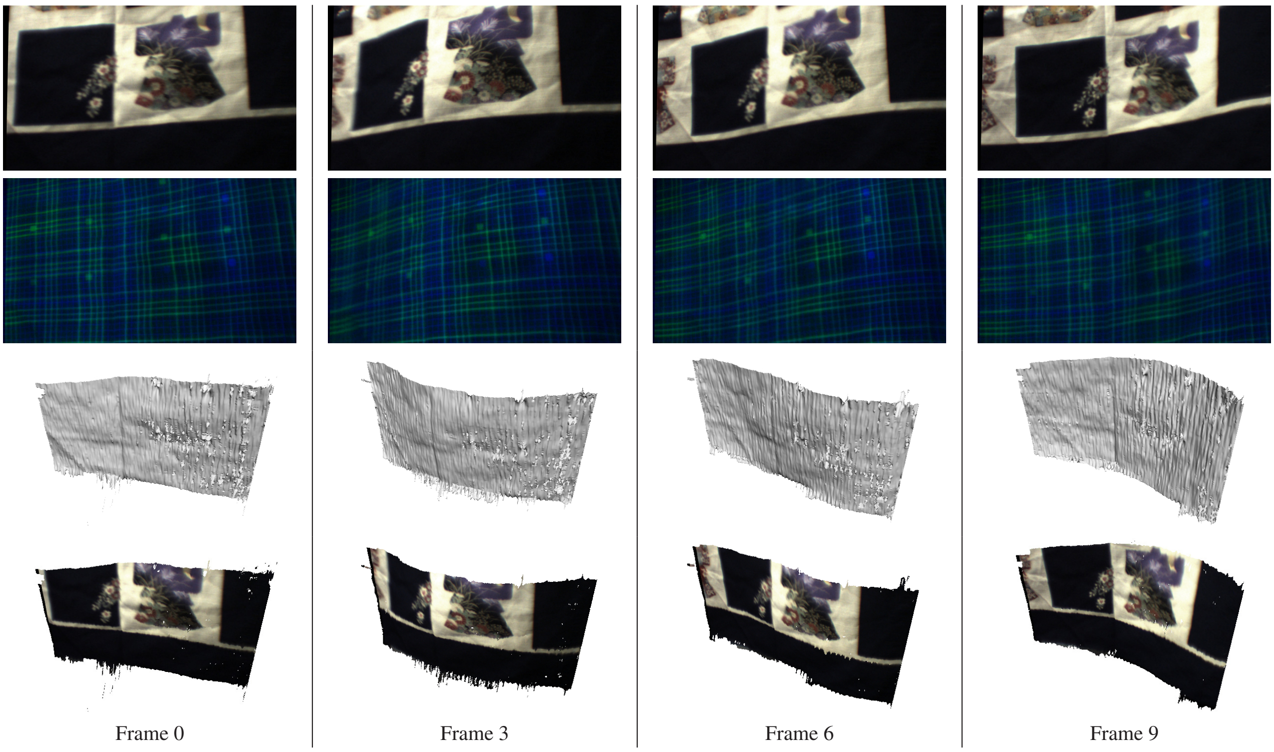


Fig. 16 The texture and shape of a waving cloth is acquired. The top row shows texture images of visible light. The second row shows the images of IR structured light. The third and fourth rows show the reconstructed shape without/with texture, respectively.

age of visible light, the IR image, the shape without texture, and the textured 3D shape, respectively. As described above, since the proposed method is based on light sectioning by using the grid pattern, the smooth surfaces are successfully reconstructed by one-shot scanning

Finally, we acquired the textured shapes of moving objects. **Figure 16** shows the input images and results of capturing a waving cloth. The images are captured at 30 frames/second. The top row shows texture images of visible light at each frame. The second row shows the images of IR structured light. The third row shows the results reconstructed without texture, and the fourth row show the shapes with texture mapping. The proposed system succeeded in capturing the shape of the moving object with texture mapping. If an object reflects IR light, the system can obtain the shape of black object in visible wavelength as shown in the result.

6. Conclusion

This paper proposes a projector-camera system that acquires textured 3D shapes based on one-shot 3D reconstruction. Since the proposed method obtain texture and shape at the same time, it can capture the shape with texture of moving objects. To accomplish simultaneous capturing texture and shape, we developed a system consisting of a projector of infrared structured light and a multi-band camera. Since the multi-band camera can observe IR and visible lights from the same viewpoint, we can accurately map the texture without affected by the error of 3D reconstruction. The projector proposed in this paper casts structured light coded by two IR wavelengths. Since the pattern is composed simply by lines perpendicular each other, the pattern is realized by multiple-lens system, which contributes to improve the efficiency of light power. Additionally, efficient calibration technique of projector-camera system by using the fixed line pattern of one-shot 3D reconstruction is proposed. With the proposed technique, only a few manual step is required for calibration, which contributes to simplify the system. In the experiments, we tested the accuracy of the proposed system and showed that it can obtain textured 3D shapes of moving objects. In future, we plan to brush up the optical design and improve the quality of the results.

Acknowledgments This work was supported in part by Strategic Information and Communications R&D Promotion Programme (SCOPE) No.101710002 (Ministry of Internal Affairs and Communications, Japan) and Funding Program for Next Generation World-Leading Researchers No.LR030 (Cabinet Office, Government Of Japan) in Japan.

References

[1] Sagawa, R., Ota, Y., Yagi, Y., Furukawa, R., Asada, N. and Kawasaki, H.: Dense 3D reconstruction method using a single pattern for fast moving object, *ICCV*, pp.1779–1786 (2009).
[2] Narayanan, P., Rander, P. and Kanade, T.: Constructing Virtual Worlds Using Dense Stereo, *Proc. 6th IEEE International Conference on Computer Vision* (*ICCV'98*), pp.3–10 (1998).
[3] Matsuyama, T., Wu, X., Takai, T. and Wada, T.: Real-Time Dynamic 3D Object Shape Reconstruction and High-Fidelity Texture Mapping for 3D Video, *IEEE Trans. Circuits and Systems for Video Technology*, Vol.CSVT-14, No.3, pp.357–369 (2004).
[4] Canesta, Inc.: CanestaVision EP Development Kit (2010), available from ⟨http://www.canesta.com/devkit.htm⟩.
[5] Mesa Imaging AG.: SwissRanger SR-4000 (2011), available from ⟨http://www.swissranger .ch/index.php⟩.
[6] Rusinkiewicz, S., Hall-Holt, O. and Levoy, M.: Real-Time 3D Model Acquisition, *Proc. SIGGRAPH*, pp.438–446 (2002).
[7] Koninckx, T. and Gool, L.V.: Real-Time Range Acquisition by Adaptive Structured Light, *IEEE Transaction on Pattern Analysis Machine Intelligence*, Vol.28, No.3, pp.432–445 (2006).
[8] Zhang, L., Curless, B. and Seitz, S.: Rapid Shape Acquisition Using Color Structured Light and Multi-Pass Dynamic Programming, *3DPVT*, pp.24–36 (2002).
[9] Je, C., Lee, S. and Park, R.: High-Contrast Color Stripe Pattern for Rapid Structured-Light Range Imaging, *Proc. Eighth European Conference Computer Vision*, pp.95–107 (2004).
[10] Zhang, S. and Yau, S.-T.: Simultaneous three-dimensional geometry and color texture acquisition using a single-chip color camera, *Optical Engineering*, Vol.47, No.12, p.123604 (2008).
[11] Waschbusch, M., Wurmlin, S., Cotting, D., Sadlo, F. and Gross, M.: Scalable 3D Video of Dynamic Scenes, *The Visual Computer* (*Proceedings of Pacific Graphics 2005*), pp.629–638, Springer (2005).
[12] Raskar, R., Welch, G., Cutts, M., Lake, A., Stesin, L. and Fuchs, H.: The Office of the Future: A Unified Approach to Image-Based Modeling and Spatially Immersive Displays, *Proc. ACM SIGGRAPH* (1998).
[13] Weise, T., Leibe, B. and Gool, L.V.: Fast 3D Scanning with Automatic Motion Compensation, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.1–8 (2007).
[14] Zhang, Z., Zhang, D., Peng, X. and Hu, X.: Color texture extraction from fringe image based on full-field projection, *Optical Engineering*, Vol.42, pp.1935–1939 (online), DOI: 10.1117/12.527563 (2003).
[15] Pan, J., Huang, P.S. and Chiang, F.-P.: Color-coded binary fringe projection technique for 3-D shape measurement, *Optical Engineering*, Vol.44, pp.3606–+ (online), DOI: 10.1117/12.631984 (2005).
[16] Zhang, S. and Huang, P.: High-resolution, Real-time 3D Shape Acquisition, *Proc. Conference on Computer Vision and Pattern Recognition Workshop*, p.28 (2004).
[17] Frueh, C. and Zakhor, A.: Capturing 21/2D Depth and Texture of Time-Varying Scenes Using Structured Infrared Light, *Proc. 5th International Conference on 3-D Digital Imaging and Modeling*, pp.318–325 (2005).
[18] Hiura, S., Yamaguchi, A., Sato, K. and Inokuchi, S.: Real-Time Object Tracking by Rotating Range Sensor, *Proc. 13th International Conference on Pattern Recognition*, Vol.1, pp.825–829 (1996).
[19] Microsoft: Xbox 360 Kinect (2010), available from ⟨http://www.xbox.com/en-US/kinect⟩.
[20] Furukawa, R., Kawasaki, H., Sagawa, R. and Yagi, Y.: 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).
[21] Kawasaki, H., Furukawa, R., Sagawa, R. and Yagi, Y.: Dynamic scene shape reconstruction using a single structured light pattern, *CVPR*, pp.1–8 (2008).
[22] Akasaka, K., Sagawa, R. and Yagi, Y.: A Sensor for Simultaneously Capturing Texture and Shape by Projecting Structured Infrared Light, *Proc. 6th International Conference on 3-D Digital Imaging and Modeling*, Montreal, Canada, pp.375–381 (2007).
[23] Zhang, Z.: A flexible new technique for camera calibration, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.22, No.11, pp.1330–1334 (2000).
[24] Snavely, N., Seitz, S.M. and Szeliski, R.: Photo Tourism: Exploring image collections in 3D, *ACM SIGGRAPH*, pp.835–846 (2006).
[25] Cignoni, P., Rocchini, C. and Scopigno, R.: Metro: Measuring error on simplified surfaces, *Computer Graphics Forum*, Vol.17, No.2, pp.167–174 (1998).

Kazuhiro Sakashita is a researcher at Service Robotics Research Group, Intelligent Systems Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Japan. He received a B.E. in Information Science from Kyoto University in 1998. He received a M.E. in Information Engineering in 2000 and a 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). He is a student member of IPSJ.

Ryusuke Sagawa is a researcher at Service Robotics Research Group, Intelligent Systems Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Japan. He received a B.E. in Information Science from Kyoto University in 1998. He received a M.E. in Information Engineering in 2000 and a 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 a lecturer of Faculty of Information Sciences, Hiroshima City University, Hiroshima, Japan. He received his M.E. and Ph.D. from Nara Institute of Science and Technology in 1993 and 1995 respectively. He became a research associate at Hiroshima City University in 1995, a lecturer at the same University in 2007. His research area includes shape-capturing, 3D modeling, appearance sampling and image-based rendering. He has won several academic awards including Songde Ma Outstanding Paper Award (best paper for ACCV) in 2007. He is a member of IPSJ, IEICE, and IEEE.

Hiroshi Kawasaki is a professor of Department of Information and Biomedial Engineering at Kagoshima University, Japan. He received his M.E. on Information Engineering in 2000 and his Ph.D. degree on Information and Communication Engineering in 2003 from the University of Tokyo, Japan, respectively. He started working at Saitama University, Japan, in 2003. Prior to Saitama University, he worked at Microsoft Research Redmond, WA, USA in 2000. He also researched at INRIA Rhone-alpes, France, as a visiting professor in 2009. His current research topic is on capturing shapes and textures of moving object. He also interested in photo-realistic rendering for VR/MR and ITS systems. He published over 100 research papers including ICCV, CVPR, IJCV, EG, MVA and ITSJ in computer vision and graphics communities. He also won several awards including Songde Ma Outstanding Paper Award (best paper for ACCV) in 2007, best paper award on PSIVT in 2009 and Nagao Prize (best paper on MIRU) in 2011. He is a member of IPSJ, IEICE, and IEEE.

Yasushi Yagi is a professor of the Institute of Scientific and Industrial Research (ISIR), Osaka university, Ibaraki, Japan. He received his Ph.D. degree from Osaka University in 1991. In 1985, he joined the Product Development Laboratory, Mitsubishi Electric Corporation, where he worked on robotics and inspections. He became a research associate in 1990, a lecturer in 1993, an associate professorin 1996, and a professor in 2003 at Osaka University. In 2012, he was chosen to be the director of the ISIR. International conferences for which he has served as chair include: FG1998 (FinancialChair), OMINVIS2003 (Organizing chair), ROBIO2006 (Program co-chair), ACCV (2007 Program chair, 2009 General Chair), PSVIT2009 (Financialchair), ICRA2009 (Technical Visit Chair), ACPR (2011 Program co-chair, 2013 General chair). He has also served as the Editor of IEEE ICRA ConferenceEditorial Board (2007–2011). He is the member of Editorial Board of International Journal of Computer Vision, the Editor-in-Chief of IPSJ Transactions on Computer Vision Applications and the Financial Chair of Asian Federation of Computer VisionSocieties. He was awarded ACM VRST2003 Honorable Mention Award, IEEEROBIO2006 Finalist of T.J. Tan Best Paper in Robotics, IEEE ICRA2008 Finalistfor Best Vision Paper, PSIVT2010 Best Paper Award and MIRU2008 NagaoAward. His research interests arecomputer vision, medical engineering and robotics. He is a fellow of IPSJ and a member of IEICE, RSJ, and IEEE.

(Communicated by *Yoshimitsu Aoki*)