Grid-based Active Stereo with Single-colored Wave Pattern for Dense One-shot 3D Scan

Ryusuke Sagawa, Kazuhiro Sakashita, Nozomu Kasuya
AIST
Tsukuba, Japan
{ryusuke.sagawa, kazuhiro-sakashita, nozomu.kasuya}@aist.go.jp

Hiroshi Kawasaki
Kagoshima University
Kagoshima, Japan
kawasaki@ibe.kagoshima-u.ac.jp

Ryo Furukawa
Hiroshima City University
Hiroshima, Japan
ryo-f@hiroshima-cu.ac.jp

Yasushi Yagi
Osaka University
Osaka, Japan
yagi@am.sanken.osaka-u.ac.jp

Abstract

In this paper, we propose a method to reconstruct the shapes of moving objects. The proposed method is a projector-camera system that reconstructs a shape from a single image where a static pattern is cast by a projector; such a method is ideal for acquisition of moving objects at a high frame rate. The issues tackled in this paper are as follows: 1) realize one-shot 3D reconstruction with a single-colored pattern, and 2) obtain accurate shapes by finding correspondences in sub-pixel accuracy. To achieve these goals, we propose the following methods: 1) implicit encoding of projector information by a grid of wave lines, 2) grid-based stereo between projector pattern and camera images to determine unique correspondences, 3) (quasi-)pixel-wise interpolations and optimizations to reconstruct dense shapes, and 4) a single-colored pattern which contributes to simplify pattern projecting devices compared to color-coded methods. In the experiment, we show the proposed method is efficient to solve the issues above.

1. Introduction

Importance of shape capture of moving objects is rapidly increasing. For example, gaming device that can capture human motion in real-time and realize a device-free interface achieved a million seller [11]. A vision system for vehicles and mobile robots are also commonly used [4]. One critical issue of these scanners is that they cannot measure shapes as accurately and densely as existing range sensors which is for static objects. If high accuracy and resolution are realized, they should be more useful for various purposes, *e.g.*, medical application, fluid analysis and so on.

There are several methods exist for capturing moving objects, such as stereo methods or time-of-flight (TOF) methods. Especially, structured-light stereo methods are suitable for capturing moving objects and have been widely researched [11, 7, 15, 21]. Structured-light methods are usually categorized into two types: temporal-encoding methods and spatial-encoding methods. Since a spatial-encoding method just requires a single input for reconstruction (*a.k.a.* one-shot scan), it is ideal to capture moving objects with high FPS. Therefore, many researches have been involved in spatial-encoding methods [17]. However, since they require certain areas to encode information on object surfaces, the resolution tends to be low and reconstruction becomes unstable.

One of the approaches to encode information in efficient ways is to use a color code. By using multiple colors, multiple bits of information can be assigned to each pixel of the camera image. A color-based coding is suitable for spatial encoding [16, 6, 24, 15, 14]. It, however, has some limitations and problems. The surface of the target objects must sufficiently reflect each color of the pattern. And, since the RGBs of off-the-shelf video projectors have overlapped spectral distribution, errors in determining colors of pixels are inevitable. To avoid those problems, several methods are proposed for efficient spatial encoding without using colors, such as dot patterns or grid patterns. Even though, there still remain several problems, *i.e.*, ambiguities on correspondences and sparse reconstruction. In this paper, we propose a one-shot scanning method which can solve the aforementioned problems with the following methods.

Grid pattern with wave patterns: By using a wave-shaped grid pattern, the intersection points can be used as features

978-0-7695-4873-9/12 $26.00 © 2012 IEEE
DOI 10.1109/3DIMPVT.2012.41

for matching. Instead of explicitly encoding the positional information of a structured light, the proposed pattern implicitly gives information which can make the order on the candidates of corresponding points.

Grid-based stereo for projector-camera systems: The proposed system can be understood as an extension of a pixel-based stereo method for a camera-pair system to a grid-based stereo for a projector-camera system. With this method, as long as connected curves are detected on captured image, global optimization can be realized.

Pixel-based interpolation using image matching: To solve a sparse reconstruction because of grid pattern, we propose a (quasi-)pixel-wise interpolation and optimization technique based on image matching to estimate depth for all the pixels.

A single-colored static pattern: To increase the stability on the colors of the target objects, we propose a single-colored static pattern based method. Thanks to the single color, wider range of projectors can be used compared to multi-colored patterns. For example, a laser projector of single narrow-band light can be used to reduce the interferences from the ambient lights.

2. Related work

TOF systems and triangulation based methods (*e.g.*, light-sectioning method or stereo method) are widely known for active measurement systems. To capture dynamic scenes with the active systems, methods based on both approaches have been researched. In many TOF laser scanning systems, a point laser is projected and the interval timebetween the emission and the detection time is measured. Since the 3D information is obtained one point at a time, it is unsuitable for capturing a entire scene in a short period of time. To capture dynamic scenes, some TOF devices project temporally-modulated light patterns and acquire a depth image at once by capturing the reflections with many detectors on a 2D image sensor [2, 10]. However, the present systems are easily disturbed by other light sources and the resolution is lower than normal cameras.

With regard to triangulation based methods, many methods of this category use point or line lasers and a scene is scanned by sweeping the lights. This type is unsuitable for dynamic scenes, because sweeping takes a time. Using area light sources is a simple solution to reduce time for scan. However, unlike the point or line light sources, ambiguity on correspondences remains. To solve it, typically two solutions are known, *i.e.*, temporal-encoding or spatial-encoding method[16].

In a temporal-encoding method, multiple patterns of illuminations are projected, and the correspondence information is encoded in the temporal modulations. Thus, it is essentially unsuitable for acquiring dynamic scenes. However, some methods are proposed to resolve this problem; *e.g.*, changing the projected patterns with high frequencies [13], reducing the required number of patterns by using phase patterns[23], or using DMD patterns[12]. Another approach is a space-time stereo, where multiple cameras are used with temporally varying illumination [25]. Although it is reported that some works can capture around 100 FPS by combining motion estimation, since these methods require multiple frames, the results degrade if object moves fast.

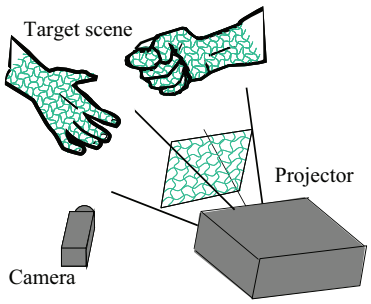


Figure 1. Scanning system: multiple wave lines are projected and their intersections are detected and used for reconstruction.

A spatial-encoding method uses a static pattern and usually requires just a single image, and thus, it is suitable to capture dynamic scenes. However, since information should be encoded in certain areas of the pattern, the resolution tends to be low. Moreover, correspondences are not stably determined because the patterns are distorted due to the color or the shapes of the object surface. Many methods have been proposed to solve the problems; *e.g.*, using multiple lines with globally-unique color combinations [20, 26], dotted lines with unique modulations of dots [9, 1], 2D area information for encoding [22, 11], using the phase of a fringe pattern [19, 18], or connections of grid patterns [8, 7, 15, 21]. However, no method has achieved sufficient performances in all aspects of precision, resolution, and stability.

In the paper, we propose a new spatial-encoding method using grid pattern to solve the aforementioned problems. Although it is known that a grid pattern based technique inevitably has ambiguity on correspondences and produces erroneous reconstruction by wrong curve detection [7, 15, 21], both problems are efficiently solved by our technique.

3. Overview

Our system consists of a single projector and a camera as shown in Fig.1. The projector casts a static pattern which is shown in Fig.3(b). The pattern is configured with vertical and horizontal sinusoidal curves to create grid shape (details are described in Sec.4). Since the pattern is static with single color, no synchronization is required, high FPS scanning is possible.

Overview of our algorithm is shown in Fig.2. First, we detect curves from a captured image. We use the curve detection method using belief propagation method proposed by Sagawa *et al*. [15]. With the method, vertical and hori-

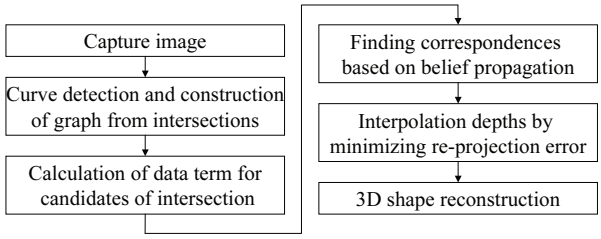
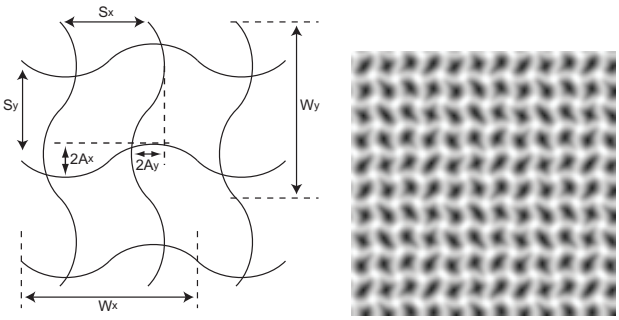
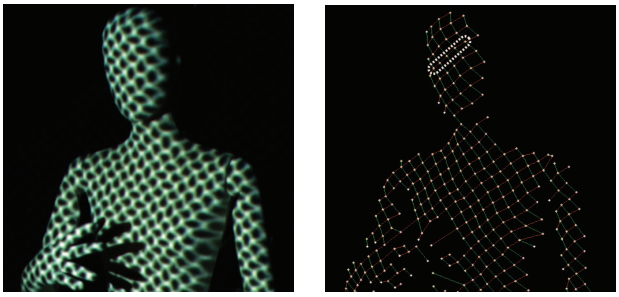


Figure 2. Algorithm overview



(a) (b)

Figure 3. Parameters of wave grid are shown in (a). S_x and S_y are the intervals between adjacent wave lines, W_x and W_y are the wavelengths of a wave line, A_x and A_y are the amplitudes of waves with respect to vertical and horizontal lines, respectively. (b) is an example of wave grid.



(a) (b)

Figure 4. Wave lines of two directions are detected separately: (a) input image, (b) detected grid. The points inside the dotted circle are used for computation in Fig.6.

zontal lines are robustly detected from a grid pattern from a single color. From the detected curves, intersection points are calculated, and a graph is also constructed by using intersection points as nodes of the graph. Then, for each intersection point, epipolar line on the projected pattern is calculated to find a correspondence. Since multiple candidates of correspondences are usually found, one solution can be determined by our belief propagation based technique. Finally, the depths for all the pixels are interpolated by matching between the pattern and the captured image and 3D shapes are densely reconstructed.

4. Implicit encoding by single-colored wave grid

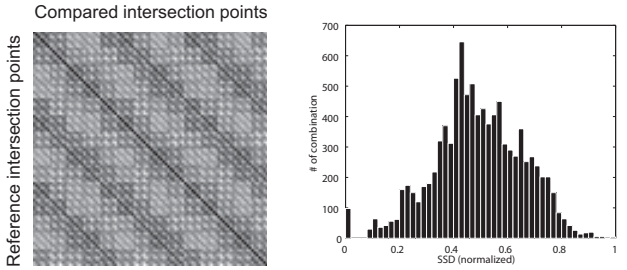
Geometrical attributes of the projectors and the cameras are usually considered to be the same. In this work, the pattern on the projector's image plane is called a projector image, and the observed image on the camera's image plane is called a camera image.

To obtain unique correspondences between the camera and projector images by spatial encoding, a complicated pattern of large window size have been required in previous methods. Moreover, while the wider baseline is desirable to improve accuracy, the observed pattern will be more distorted, which makes it difficult to decode the pattern in practical cases. Therefore, we use a simple but informative pattern that is easy to detect and decode.

In this paper, we propose a pattern that gives information which can make the order on the candidates of corresponding points rather than get the unique correspondence through decoding process. The proposed pattern consists of vertical and horizontal directions of wave lines, which forms a grid pattern. Because each wave line is simple, it is easy to detect curves, and the position of a curve can be calculated in sub-pixel accuracy by detecting peaks of intensities of the curve.

The wave line is a sinusoidal pattern, which is periodic and self-recurring. The grid of wave lines, however, can give information for finding correspondences. The proposed method uses the intersection points of vertical and horizontal wave lines as feature points. The arrangement of intersection points is determined by the intervals and the wavelength of the wave lines. In the paper, we use the same interval and wavelength for all the vertical and horizontal wave lines. However, as described in the following, because the interval of the vertical wave lines is not equal to the integral multiple of the horizontal wavelength, the intersection points appear at the different phases on the wave pattern; it means that the local pattern around an intersection point has local uniqueness, and it can be used as a discriminative feature. In this paper, we also use 'wave patterns' to refer to the wave lines.

The local pattern around a intersection point is not globally-unique in the whole pattern and periodic. Therefore, the same pattern occurs at every N_x and N_y wave lines along the horizontal and vertical axes, where $N_x = \mathrm{lcm}(S_x, W_x)/S_x, N_y = \mathrm{lcm}(S_y, W_y)/S_y$ where $\mathrm{lcm}(a, b)$ is the least common multiple of a and b. Hereafter, subscript letter x means the symbol describes values about horizontal axis, and y about vertical axis. S_x and S_y are the intervals between adjacent wave lines, and W_x and W_y are the wavelengths, as shown in Fig.3(a). The patterns, however, can be discriminative in each cycle. Fig.3(b) shows an example with $S_x = 10, S_y = 11, W_x = W_y = 14, A_x = A_y = 1$ (pixels), where A_x and A_y are the amplitudes of waves. In



(a) (b)

Figure 5. (a) The matrix of SSD values between all pairs of intersection points. Each pixel represents the SSD value of a pair. Bright pixels indicate large differences. (b) The histogram shows the distribution of SSD values.

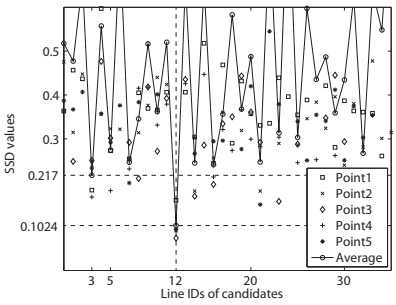


Figure 6. The SSD values are calculated for five points on the same wave line in Fig.4(b). These points are compared to the multiple candidates in the projector image. Each marker corresponds to the one of five points. The black line is the averaged value of five points. The correct ID of the wave line is 12 in this case.

this case, each cycle has 7 and 14 wave lines along horizontal and vertical axes, respectively. Consequently, 98 ($= 7 \times 14$) intersection points of different patterns exist in the rectangle. Fig.5(a) shows the matrix of sum of squared differences (SSD) between all pairs of intersection points, where the window size is 10×10 pixels. A pixel in the matrix represents the SSD value of a pair. Bright pixels indicate large differences. Since the dark pixels exist sparsely, the number of similar patterns is minor, which can be seen from the histogram of the distribution of SSD values shown in Fig.5(b). This indicates the local pattern around intersection points can be used as discriminative features.

In stereo matching, the candidates of corresponding points are restricted to the points on the epipolar line. If a intersection point is located within a certain distance from the epipolar line, it is chosen as a candidate. The number of candidates depends on the position of intersection points in the camera image. Since the candidates are sparsely located in the projector image, the number of candidates is much smaller than the case of usual pixel-based stereo without restricting the search range.

Fig.6 shows the SSD values of the five points on the same wave line in Fig.4(b). These points are compared to the candidates in the projector image. Each marker in the figure corresponds to the one of five points. Since the points are

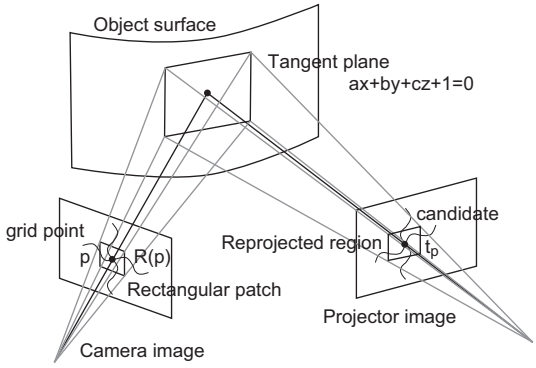


Figure 7. The rectangular patch around a grid point p is reprojected onto the projector's image plane. t_p is one of candidates of correspondence for p.

on the same wave line, the correct ID of wave line is the same, which is 12 in this case. The SSD values of each points have the minimum value at ID 12, but the difference between the value at ID 12 and others is not apparent, since the feature by wave lines are not so strong. The averaged value indicated by the black line, however, clearly has the minimum at the ID 12; the value at ID 12 is 0.102, while the secondary minimum is 0.217. It shows the SSD values at intersection points can be used as the feature to compare, and the connectivity given by wave lines can be used as powerful constraint for regularization.

5. Projector-camera stereo on grid

To find the best combinations of correspondences, a new optimizing method using regularization with matching cost of local patterns is introduced. The proposed method uses the grid of wave patterns detected by curve detection shown in Fig.4(b). An intersection point of wave grid in the camera image is hereafter called a grid point. If grid points are connected with each other by the grid, they should be on the same wave line on the projector image, which can be used as regularization. It is assumed that the connectivity of grid points is given by the line detection. The connectivity is however sometimes wrong, typically because of the occluding boundaries. Therefore, we propose a method to solve the problem to assign the corresponding point for each grid point using the energy minimization on the grid.

5.1. Data term for energy minimization

First, we calculate the matching costs for all the candidates as the data term for energy minimization. The cost is computed by SSD between the captured image and the projector image. Though, since the position of a grid point has some error and the pattern observed by the camera is distorted by the surface geometries, the simple SSD with rectangular patch is unsuitable for the data term. Therefore, we use the tangent plane of patch around the grid point to calculate a better matching cost and determine the correspondence with each candidate in sub-pixel accuracy.

We assume that the surface around a grid point is locally planar as shown in Fig.7, and the 3D plane is represented by $ax + by + cz + 1 = 0$, where a, b and c are parameters of a plane. We estimate the parameters that accounts for the distortion by minimizing the SSD. The procedure is as follows.

1. Project the rectangular patch around a grid point in the camera image onto the 3D plane, and re-project it onto the projector's image plane.
2. Calculate the SSD between the intensities of the projector image and the re-projected patch.
3. Modify a, b and c to minimize the SSD values.
4. Iterate the above steps for several times.

The initial values of a, b and c are set so that the plane contains the point computed by triangulation and parallel to the camera's image plane. The SSD value is calculated by the following equation.

$$SSD_{a,b,c}(p) = \sum_{p' \in R(p)} (I_c(p') - I_p(H_{a,b,c}(p'))^2, \quad (1)$$

where $R(p)$ is the rectangular patch around p and $H_{a,b,c}(p')$ is the transformation by re-projection of p' onto the projector image. $I_c(\cdot)$ and $I_p(\cdot)$ are the intensities of the camera and projector images, respectively.

5.2. Finding correspondences by energy minimization on grid

Now the grid consists of grid points $p \in V$ and the connections $(p, q) \in U$, where p and q are grid points, V is the set of grid points, and U is the set of edges of the grid graph. A grid point p has the candidates of corresponding points $t_p \in T_p$ in the projector image, where T_p is the set of candidates for the grid point p. We define the energy to find correspondences as follows:

$$E(T) = \sum_{p \in V} D_p(t_p) + \sum_{(p,q) \in U} W_{pq}(t_p, t_q), \quad (2)$$

where $T = \{t_p | p \in V\}$. $D_p(t_p)$ is the data term of assigning a candidate t_p to p. $W_{pq}(t_p, t_q)$ is the regularization term of assigning candidates t_p and t_q to neighboring grid points.

The data term is the SSD calculated by the method described in previous section. The regularization term is defined as follows

$$W_{pq}(t_p, t_q) = \begin{cases} 0 & t_p \text{ and } t_q \text{ are on the same wave line} \\ \lambda & \text{otherwise,} \end{cases} \quad (3)$$

where λ is a user-defined constant. The energy is minimized based on belief propagation [5] in this paper.

An advantage of using energy minimization to enforce structures is that they can be "soft constraints." This is important because there is always a chance that erroneous grid connections occur in actual case. With our method, while wrong connection was determined and removed at the line detection before 3D reconstruction in the previous methods [15], wrong connection removal and 3D reconstruction are simultaneously accomplished to achieve dense and better result. This is because not only local features by wave pattern but also epipolar constraint are used for removing wrong connection, while only color information was used in [15].

6. Generating dense shape by interpolating grid points

The correspondences for sparse grid points are obtained by the grid-based stereo. The next step is to obtain dense correspondences by using all the pixels. We first calculate depth values of densely resampled pixels by interpolating the grid points using estimated local planes of surrounding grid points for each pixel. Then, the densely resampled depth values are optimized by minimizing the difference of intensity for all the pixels between camera image and projector image. Although Sagawa *et al.* [14] also achieved dense reconstruction, it was based on interpolation between grid lines, whereas, in this work, independent depth estimation for each pixel is achieved by (quasi-)pixel-wise optimization based on photo-consistency.

6.1. Calculating depths for resampled pixels

If the viewing vector of a resampled pixel x from the camera origin is $(u, v, 1)$, the depth d_x for the pixel is computed by

$$d_x = \frac{-1}{a_x u + b_x v + c_x}, \quad (4)$$

where a_x, b_x and c_x are the parameters computed for the pixel. a_x for each pixel is interpolated as follows:

$$a_x = \frac{\sum_p G(|p - x|) a_p}{\sum_p G(|p - x|)}, \quad (5)$$

where p is a grid point, $G(\cdot)$ is a Gaussian function, and $|p - x|$ is the distance between p and x. b_x and c_x are calculated in similar manner by weighted averaging.

6.2. Optimizing depths for all resampled pixels

We resample the image pixels to generate a dense mesh model, for example, every three pixels for both vertical and horizontal directions. The initial position of each vertex is given by using the depth calculated by Eq.(4). The generated mesh model consists of triangles by connecting the adjacent resampled pixels as shown in Fig.8. To optimize the depths iteratively, we parameterize d_x by using small movement Δd_x. The depth of pixel x in Fig.8 is calculated as follows.

$$d_x + \Delta d_x = [1 - w_{x_2} - w_{x_3}, w_{x_2}, w_{x_3}] \begin{bmatrix} d_{x_1} + \Delta d_{x_1} \\ d_{x_2} + \Delta d_{x_2} \\ d_{x_3} + \Delta d_{x_3} \end{bmatrix}, \quad (6)$$

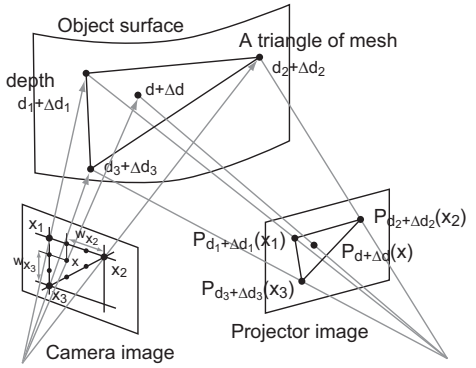


Figure 8. x_1, x_2 and x_3 are resampled pixels. Each pixel is reprojected onto the projector's image plane by using the depth. The transformation is represented by $P_{D+\Delta D}(x)$, while each reprojection uses a subset of $D + \Delta D$ for transformation.

where w_{x_2} and w_{x_3} are the weights for linear interpolation. $d_x + \Delta d_x$ can be calculated for all pixels in the camera image by interpolating the depths of resampled pixels.

Now, $D + \Delta D$ is a vector of $d_x + \Delta d_x$ by collecting depths of all the vertices. The reprojection error compared with the projector image is calculated by using all the pixels as follows:

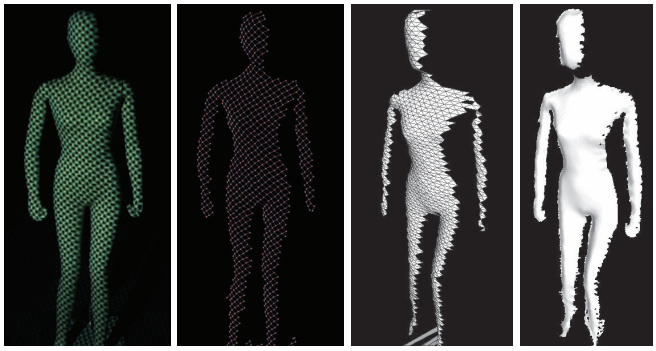
$$E(\Delta D) = \sum_x (I_c(x) - I_p(P_{D+\Delta D}(x))^2 + \gamma \sum_{x,x'} (\Delta d_x - \Delta d_{x'})^2, \quad (7)$$

where the reprojection onto the projector image is represented by $P_{D+\Delta D}(x)$, while each reprojection uses a subset of $D + \Delta D$ for transformation. x and x' are adjacent vertices. γ is a user-defined parameter for regularization. We estimate the parameter ΔD by minimizing the error, and iterate the reprojection and minimization alternatively until convergence by updating the depths D. Finally, the optimized position of each vertex is calculated by the updated depths D.

7. Experiments

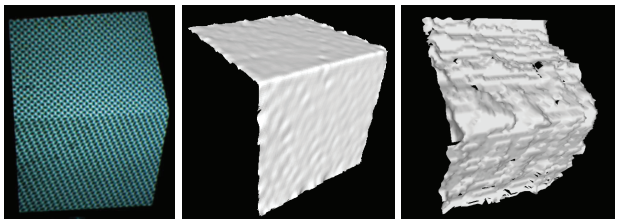
We used a camera of 1600×1200 pixels and a projector of 1024×768 pixels. The image sequences were captured at 30FPS. We used a PC with Intel Core i7 2.93GHz and NVIDIA GeForce 580GTX. The algorithms of image processing and reconstruction are implemented with CUDA.

First, we show the effectiveness of the proposed wave pattern by comparing with a straight-line pattern. Fig.9 is the result of 3D reconstruction with the proposed wave pattern shown in Fig.3(b). Fig.9(a) is the input image and (b) is an obtained grid pattern after the projector-camera stereo described in Sec.5 is applied. The grid lines near the occluding boundary, *i.e.*, head and neck of the mannequin, were successfully disconnected at the stereo matching process. Fig.9(c) shows the result of 3D reconstruction with our method. The number of the grid points was 943 and the



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

Figure 9. 3D reconstruction with the proposed wave pattern: (a) input image, (b) detected grid after stereo matching, (c) reconstruction results of the grid, (d) reconstruction after interpolation.



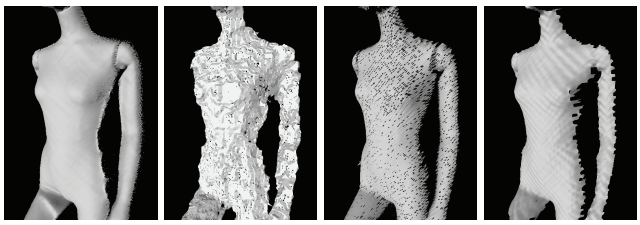
(a) (b) (c)

Figure 10. 3D reconstruction of a cube-shaped object: (a) input image, (b) the result of 3D reconstruction by the proposed method, (c) the result by Kinect.

average number of candidates of corresponding point for each grid point was 41. The computational time for curve detection and the grid-based stereo were 64.4 msec, and 27.5 msec, respectively. Although the search space was full viewing volume, thanks to the sparsity of the grid pattern, the computational cost is still low. The proposed method generated the dense shape as shown in Fig.9(d) by the interpolation method described in Sec.6. The number of vertices of the 3D model was 24,618. The number of iteration for optimization was five, and the computational time for interpolation was 117.5 msec. The total time including curve detection was 209.5 msec.

Therefore, the proposed method realizes on-line processing with additional processing including visualization, at about 10 frames/second with the grid-based stereo. The interpolated shape can be generated at 3 frames/second without any pipeline processing. Since the implementation is not fully optimized, the frame-rate can be improved and it is our future work.

Next, the accuracy of the proposed method was evaluated by capturing a cube-shaped object as shown in Fig.10. The size of the cube was 0.2m square and the distance from the camera was about 1.0m. Each face of the reconstructed cube was fit to a plane to calculate RMSE. The average of RMSE of two planes was 0.36mm. Fig.10(c) is the result obtained by Microsoft Kinect, of which the RMSE of fitting a plane to each face is 1.78mm. The errors of the pro-



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

Figure 11. Comparison of different methods: (a) the temporal-encoding method by projecting phase-shift pattern, (b) Kinect, (c) the spatial-encoding method by projecting color-encoded grid pattern [14], and (d) the proposed method.

posed method is smaller than Kinect, and can be considered sufficiently small for practical usage.

Fig.11 shows another comparison of the result by different methods: (a) the temporal-encoding method by projecting phase-shift pattern, (b) Kinect, (c) the spatial-encoding method by projecting color-encoded grid pattern [14], and (d) the proposed method. Since the temporal-encoding method have advantage in terms of accuracy, we used it as the ground truth for quantitative comparison. The difference between meshes are calculated by using a method proposed by Cignoni [3]. The RMSEs from (a) are (b) 22.9mm, (c) 4.9mm, and (d) 3.9mm, respectively. The distance from cameras to the object is about 2.1m, and the baseline between camera and projector for the cases of (c) and (d) is 0.5m. The propose method (d) gave the best performance. The error is caused by the artifact of projected lines. Since the proposed method minimizes the difference between camera and projector images, the optimization is affected by optical condition including lens focus, surface characteristics, etc. Therefore, the artifact occurs in some conditions. To reduce this artifact is one of our future work. The error of (c) is slightly larger than the proposed method. The error is mainly caused by wrong calculation of phase to interpolate lines, which results in small holes of the mesh. The error by Kinect is larger than the other two. It is considered mainly because it has narrow baseline and wide field of view compared to the other cases.

Finally, we show the results of capturing moving objects. Fig.12 and Fig.13 show the results of capturing the scene of opening a hand and punching human, respectively. Since the proposed method is one of one-shot method, it can generate 3D shapes even if the target moves in fast motion.

8. Conclusion

In this paper, efficient dense 3D reconstruction method from a single image (one-shot scan) using single-colored static pattern projector is proposed. The method is unlike previous one-shot scanning methods and utilizes implicit encoding technique which consists of grid pattern with sinusoidal curves. With the pattern, irregularity of pattern is increased and solution becomes stable. We also propose a new approach to extend stereo technique to projector and camera system by using connectivity of grid pattern. At the final reconstruction step, a pixel based interpolation method which can recover dense 3D shape is presented. In the experiments, we evaluated the accuracy of our method compared to state-of-the-art one-shot scan technique and proves strength of our method. In addition, successful 3D reconstruction of moving objects is demonstrated. In the future, we plan to use multiple devices to reconstruct entire shape of moving object.

Acknowledgment

This work was supported in part by SCOPE No.101710002 and NEXT program No.LR030 in Japan.

References

[1] Artec. United States Patent Application 2009005924, 2007j. 2

[2] Canesta, Inc. CanestaVision EP Development Kit, 2010. http://www.canesta.com/devkit.htm. 2

[3] P. Cignoni, C. Rocchini, and R. Scopigno. Metro: measuring error on simplified surfaces. *Computer Graphics Forum*, 17(2):167–174, 1998. 7

[4] DARPA. 2007 Urban Challenge, 2007. 1

[5] P. Felzenszwalb and D. Huttenlocher. Efficient belief propagation for early vision. *IJCV*, 70:41–54, 2006. 5

[6] 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. 1

[7] H. Kawasaki, R. Furukawa, , R. Sagawa, and Y. Yagi. Dynamic scene shape reconstruction using a single structured light pattern. In *CVPR*, pages 1–8, June 23-28 2008. 1, 2

[8] T. Koninckx and L. V. Gool. Real-time range acquisition by adaptive structured light. *IEEE Transaction Pattern Analysis Machine Intelligence*, 28(3):432–445, 2006. 2

[9] M. Maruyama and S. Abe. Range sensing by projecting multiple slits with random cuts. In *SPIE Optics, Illumination, and Image Sensing for Machine Vision IV*, volume 1194, pages 216–224, 1989. 2

[10] Mesa Imaging AG. SwissRanger SR-4000, 2011. http://www.swissranger .ch/index.php. 2

[11] Microsoft. Xbox 360 Kinect, 2010. http://www.xbox.com/en-US/kinect. 1, 2

[12] S. G. Narasimhan, S. J. Koppal, , and S. Yamazaki. Temporal dithering of illumination for fast active vision. In *Proc. European Conference on Computer Vision*, pages 830–844, October 2008. 2

[13] S. Rusinkiewicz, O. Hall-Holt, and M. Levoy. Real-time 3D model acquisition. In *Proc. SIGGRAPH*, pages 438–446, 2002. 2

[14] 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. 1, 5, 7

[15] 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. 1, 2, 5

[16] 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. 1, 2

[17] J. Salvi, J. Pages, and J. Batlle. Pattern codification strategies in structured light systems. *Pattern Recognition*, 37(4):827–849, 4 2004. 1

[18] G. Sansoni and E. Redaelli. A 3d vision system based on one-shot projection and phase demodulation for fast profilometry. *Meas. Sci. Technol.*, 16:1109–1118, 2005. 2

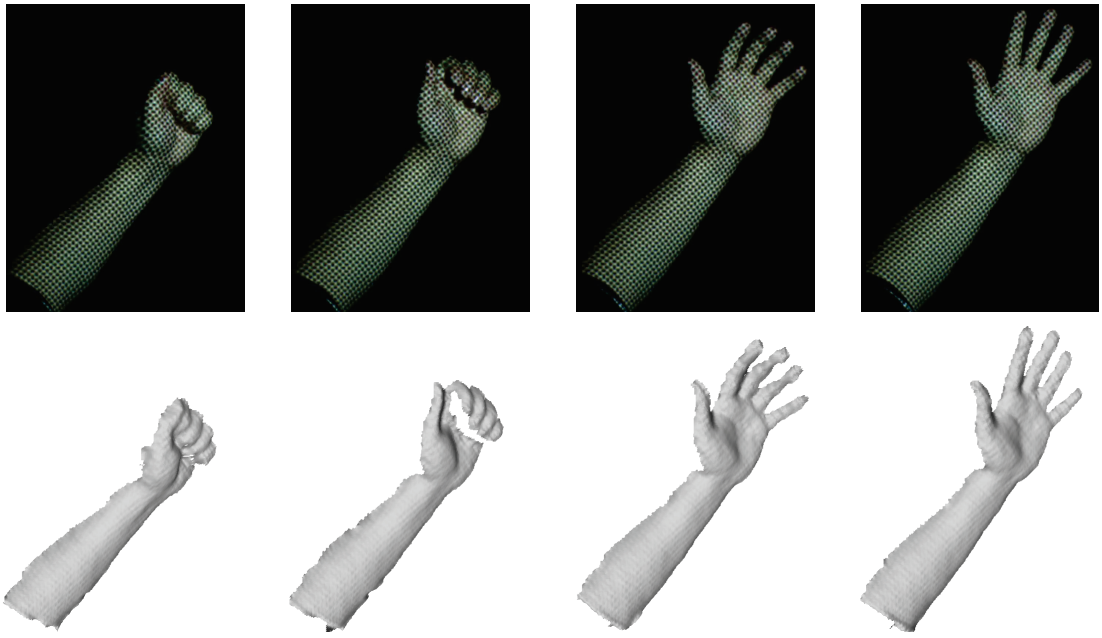


Figure 12. The 3D shapes of a hand is reconstructed during opening the fingers.

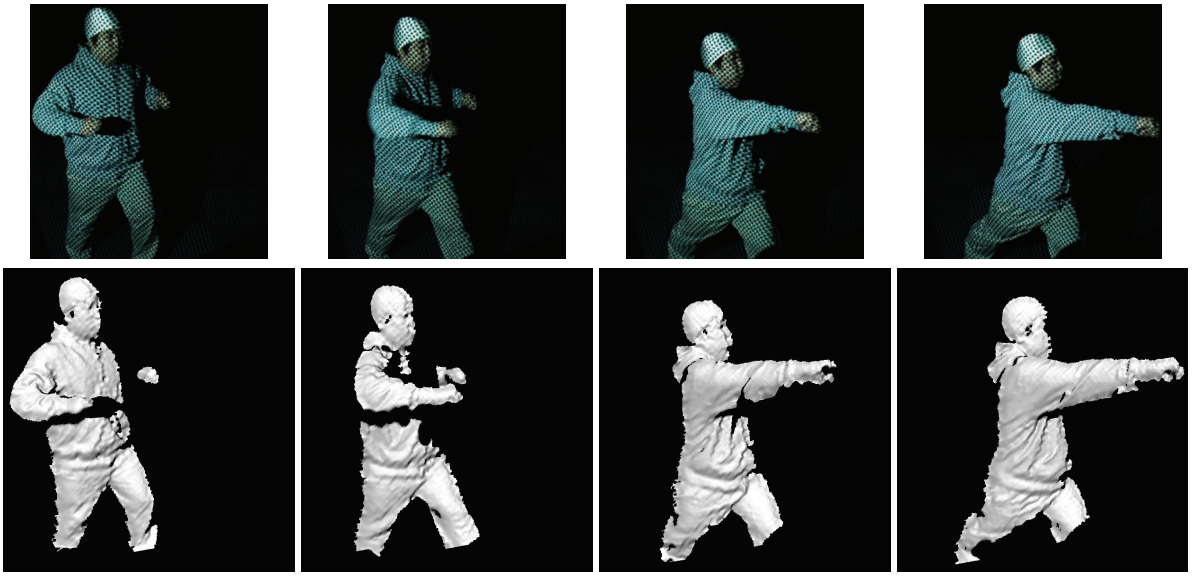


Figure 13. The 3D shapes of a punching scene are reconstructed. The detailed shape of the cloth is successfully obtained.

[19] V. Srinivasan, H. Liu, and M. Halioua. Automated phase-measuring profilometry of 3-d diffuse objects. *Applied Optics*, 23(18):3105–3108, 1984. 2

[20] J. Tajima and M. Iwakawa. 3-D data acquisition by rainbow range finder. In *ICPR*, pages 309–313, 1990. 2

[21] A. O. Ulusoy, F. Calakli, and G. Taubin. One-shot scanning using de bruijn spaced grids. In *The 7th IEEE Conf. 3DIM*, 2009. 1, 2

[22] P. Vuylsteke and A. Oosterlinck. Range image acquisition with a single binary-encoded light pattern. *IEEE Trans. on PAMI*, 12(2):148–164, 1990. 2

[23] T. Weise, B. Leibe, and L. V. Gool. Fast 3D scanning with automatic motion compensation. In *CVPR*, 2007. 2

[24] L. Zhang, B. Curless, and S. Seitz. Rapid shape acquisition using color structured light and multi-pass dynamic programming. In *3DPVT*, pages 24–36, 2002. 1

[25] 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

[26] S. Zhang and P. Huang. High-resolution, real-time 3D shape acquisition. In *Proc. Conference on Computer Vision and Pattern Recognition Workshop*, page 28, 2004. 2