Abstract

In this paper, we propose a new technique to achieve one-shot scan using single color and static pattern projector; such a method is ideal for acquisition of moving objects. Since projector-camera systems generally have uncertainties on retrieving correspondences between the captured image and the projected pattern, many solutions have been proposed. Especially for one-shot scan, which means that only a single image is required for shape reconstruction, positional information of a pixel of the projected pattern should be encoded by spatial and/or color information. Although color information is frequently used for encoding, it is severely affected by texture and material of the object and leads unstable reconstruction. In this paper, we propose a technique to solve the problem by using geometrically unique pattern only with black and white color which further considers the shape distortion by surface orientation of the shape. Our technique successfully acquires high precision one-shot scan with an actual system.

Single Color One-shot Scan using Modified Penrose Tiling Pattern

Hiroshi Kawasaki, Hitoshi Masuyama, Ryusuke Sagawa and Ryo Furukawa

August 16, 2016

1 Introduction

Methods to capture moving objects have been increasingly attracting attention these days. For example, recently, inexpensive scanning devices developed for entertainment purposes achieved over 24 million sales so far [13]. Since the purpose of the scanner is a motion capture for games to realize a device-free user interface, the importance on accuracy and density of 3D reconstruction is not high and they are lower than that of existing range sensors. If higher accuracy with denser resolution is achieved for such devices, applicable areas of the sensor can be greatly increased and many new systems will be developed, *e.g.*, medical application, robot vision, fluid analysis and so on. However, considering the state of the art of active 3D scanning techniques, the acquisition of high accuracy and density is still a challenging task.

Reviewing the current active scanning methods for capturing moving objects, the methods can be categorized into two types; a structured-light based method and a time-of-flight (TOF) method. Since the structured-light based method can be easily developed by off-the-shelf devices, such as a video projector and CCD cameras, it becomes more popular than before. Structured-light based methods can be further categorized into two types, such as temporal-encoding and spatial-encoding based methods and detailed explanations are summarized in [19]. Among them, since a spatial-encoding method only requires single image for reconstruction and is ideal for capturing moving objects, there have been many researches conducted along with this approach (the technique is also known as a one-shot scan). One critical issue on one-shot scan is its instability during reconstruction. This is because they require certain area on the pattern to encode the positional information of the projector's pixel, however, larger area is more easily affected by shape distortion of the object's surface. To overcome the pattern distortion caused by the shape of the object, since stripe patterns have a beneficial property that the order of the pattern along an epipolar line is preserved for smooth surfaces, the stripe patterns are widely researched [18, 8, 24].

One drawback of the method is that, since those methods usually use color information for identification process, they are severely affected by the texture of the object surface. To avoid such texture problem, several methods are proposed for efficient spatial encoding using only black and white color to configure the pattern with dots [13, 1] or grid pattern [9, 17, 16]. In spite of all the efforts, there still remain several problems, *i.e.*, difficulty to accurately reconstruct the shape by using short base line [13], sparse reconstruction by wide interval of grid pattern [9, 17] and the instability caused by wrong line detection [9, 17, 16].

In this paper, we propose a one-shot scanning method, which can solve the aforementioned problems with the following approach.

1) Robust and unique pattern for active stereo: To increase the stability on retrieving correspondences, we propose a unique pattern considering shape distortion caused by geometric features of the object surface. To avoid the pattern disturbance caused by the texture of the object surfaces, we make the pattern using only black and white color. Because of such limitation, the variety of possible patterns is much smaller than color based patterns. Further, since patterns are naturally distorted by the unique shape of the object surface, different patterns may become similar and indistinguishable for the actual 3D object, and thus, its accuracy is decreased. To solve such problem, we discuss the preferable feature on pattern and propose a unique pattern using modified Penrose tiling [7]. The pattern has a graph representation which consists of nodes and edges. The nodes can be used as features that can be distinguished by, for example, the order of nodes (the number of edges connected to the nodes). Since such topological information is well preserved during geometric transformation, robust correspondences with wide base-line can be acquired with a certain matching algorithm.

2) Efficient stereo algorithm for projector camera system: Topology information is ideal to increase the uniqueness of patterns, especially for shape distortion. However, stable extraction of topology information itself is not easy. In this paper, we propose a stereo technique which implicitly uses topology information rather than directly extract them from the pattern. During the experiment, the sum of squared differences (SSD) is used for the matching scores which further considers all the possibility of shape distortion; such distortion is caused by affine transformation of surface orientations. Because of the additional search space, computational cost becomes larger than usual stereo algorithm. To mitigate the huge calculation cost, we propose a coarse-to-fine strategy (pyramid approach) and affine search is only done in coarse level. Estimated surface orientation in coarse level is used in finer levels.

3) Global optimization to decrease inaccurate reconstruction: Since the technique is based on stereo, small errors frequently occur and it is known that such errors are efficiently solved by MRF based global optimization techniques. In our method, since the matching scores of each pixel are calculated by SSD considering affine transformation, it tends to have more local minima than usual stereo algorithm. Among such multiple candidates, correct correspondence is expected to be efficiently determined by using our unique pattern with global optimization technique. One problem on MRF algorithm is computational cost. Since our stereo algorithm already adopt coarse-to-fine strategy to estimate surface orientation, MRF optimization is only applied on coarse level to reduce computational time. In this paper, we use belief propagation method for MRF optimization.

4) Shape reconstruction near object boundary using edge connectivity: For window matching based approach, reconstructed shapes inevitably have noise especially near the objects' boundaries. In our method, since the pattern of our method has graph representation, we use the edge connectivity to efficiently remove only the noise, but reconstruct the correct shape near the boundary.

The paper is organized as follows. Firstly, we briefly review the related works on active stereo technique especially for moving object. Then, we explain an overview of our method followed by an introduction of our original pattern which preserves the uniqueness despite of the distortion caused by the object shape. Then, we explain the

details of our method and show the effectiveness of our method with several experiments using the real system. Finally, we conclude the paper.

2 Related works

Triangulation based methods (*e.g.*, light-sectioning method or stereo method) and time-of-flight(TOF) based methods are widely known for active measurements. Since many TOF based systems use point lasers and need to sweep the object, they are not suitable to capture the 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 using a special 2D image sensor [3]. However, the present systems are easily disturbed by other light sources and the resolution is lower than normal cameras because of its requirement of complicated high-precision sensor. Furthermore, there is a problem of ambiguity because of periodical light modulation.

In regards to triangulation based methods, many methods which use point or line lasers to scan a scene by sweeping them are proposed. These types of methods are also unsuitable for dynamic scene capturing, because sweeping takes time. Using area light source such as video projectors is a simple solution to reduce the scanning time. However, unlike the point or line light sources, there remains an ambiguity on correspondences. For the solution, typically two methods are known, *i.e.*, temporal-encoding or spatial-encoding methods [18].

In a temporal-encoding method, multiple patterns 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 by capturing with high frequencies [15, 23, 14]. Although it is reported that some works can capture around 100 FPS by combining motion compensation [23], the quality of the results is degraded if the object moves fast, because these methods still require multiple frames.

A spatial-encoding method, which uses a static pattern, usually requires just a single image, and thus, it is suitable to capture dynamic scenes. However, since the information should be encoded within certain area of the pattern, the resolution tends to be low. Moreover, correspondences are not stably determined because the patterns are distorted due to the color and the shapes of the object surface. Many methods have been proposed to solve the problems; *e.g.*, using multiple stripes with globally-unique color combinations [20, 25], dotted lines with unique modulations of the dots [12, 2], 2D area information for encoding [22, 13], or connections of grid patterns [11, 9, 17, 21]. However, no method has achieved a sufficient performance in all aspects of precision, resolution, and stability.

In this paper, we propose a simple technique to solve the aforementioned problems using a new pattern where shapes are reconstructed by window matching technique and the pattern of matching window is always unique along an epipolar line under affine transformation. With our analysis and experiments using real pattern, it shows the future direction of optimal pattern for one-shot scan with a single color.

3 System Overview

3.1 System configuration

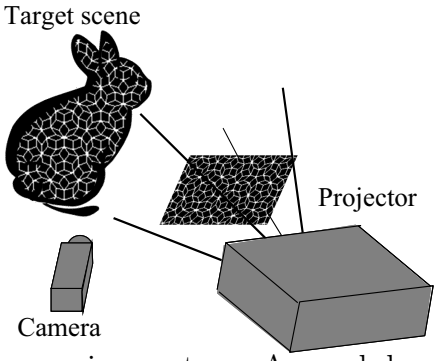


Figure 1: Overview of the scanning system: A graph-based pattern is projected from the projector and captured by the camera

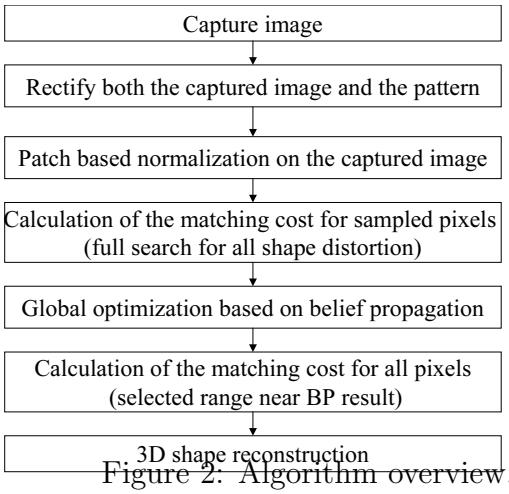


Figure 2: Algorithm overview.

Our system consists of a single projector and a camera as shown in Fig.1. The projector casts a static pattern onto the object and the scene is captured by the camera. Then, shapes are reconstructed from a single image by using the stereo algorithm. The pattern consists of lines (edges) and intersections (nodes) which make a graph representation (details are described in the next section). Since the pattern is static, no synchronization is required. We assume that the calibration parameters including intrinsic and extrinsic parameters of the camera and the projector are known.

3.2 Algorithm overview

Overview of our algorithm is shown in Fig.2. First, we rectify both the captured image and the projected pattern. Then, we normalize the captured image for better calculation of the matching cost. In this process, since the environmental lighting condition and the texture may not be uniform, adaptive normalization is applied. In the next step, matching costs are calculated for coarse level of two layers of pyramid, *e.g.*, 20 pixels interval in our experiments. In this step, matching cost for all the direction of surface normal (2D) is calculated. Using the cost, global optimization is applied using belief propagation [5] and the surface orientation is also estimated at the same time. Then, as for the fine level of the bottom layer of pyramid, matching costs for all the pixels are calculated. The search range is restricted only near the result of the higher layer. Finally, the depths for all the pixels are reconstructed using the estimated disparity by a triangulation method.

4 Robust and unique pattern for active stereo

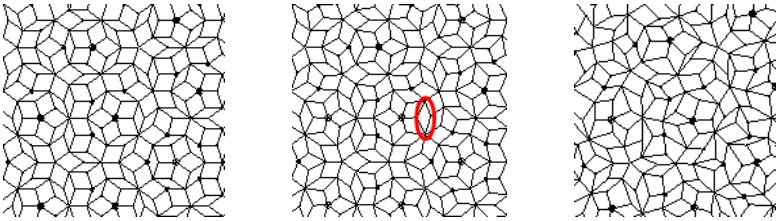
In one-shot active stereo, the pattern projected to the target is important in order to achieve sufficient performances. The pattern is projected to the target surface and observed by the camera. The observed pattern is deformed by the geometries of the surface. The local deformation of this process can be represented as 3D homographies by regarding the local surface as a small plane (a patch).

One of the patterns whose geometric property is not changed under 2D homography is a line. However, a simple line does not have much geometric feature, thus, it is not appropriate for stereo matching as it is. One of the possible solution for this is to use a pattern of a planar graph (graph that can be embedded in the plane without intersecting edges) with an appropriate geometric complexity.

For a planar graph as a pattern, one important feature is the number of edges that are connected to each of the nodes (orders of the nodes), or edge connections between the nodes. Those features are topological properties of the graph, which does not change under 2D homographies. In this experiment, we propose to use a pattern generated by Penrose tiling [7], which has plenty of such features.

Penrose tiling is a kind of tiling (filling a plane with some geometric shapes without overlaps or gaps) that can be generated by a small number of tiles. The generated pattern is known to have no translational symmetry. Among several kinds of Penrose tiling, we use patterns that are generated by two kinds of rhombuses [7] (rhombus tiling). The generation can be easily achieved by using a recursive algorithm. An example of rhombus tiling is shown in Fig.3(a).

If orders of nodes are regarded as features, a graph that includes nodes with many kinds of orders has more distinctive features. However, if the order of a node is too large (*i.e.*, too many edges are connected to the node), some of the edges may easily become indistinguishable under deformation caused by homographies, *e.g.*, compression of the texture. The orders of nodes are 3, 4, or 5 for the proposed pattern. Thus, the order is at most 5, although it can have several varieties of values. Moreover, the density of nodes in the pattern is uniform. This is useful to achieve dense reconstruction during the actual measurement. From the above-mentioned properties, the graph generated by rhombus tiling can be considered to have good properties as an active stereo pattern.



(a) (b) (c)

Figure 3: Graph pattern generated by Penrose tiling: (a) a basic pattern, (b) a basic pattern that is modified so that thin rhombuses (one of them is marked by a red ellipse) are enlarged, and (c) a pattern in which the positions of the nodes are disturbed to reduce the repetition of similar patterns.

The minimum angle of corners of the rhombuses, which is shown in Fig.3(a), is as small as 36 degrees. Such a narrow angle is inappropriate for the pattern, because the two edges of the corners may become difficult to distinguish by the homographic deformations which compress the texture. In this work, we modified positions of the nodes so that the minimum angle of corners is increased. To achieve this, we assume repulsive forces between the nodes of the graph that would be generated if all the nodes had the same electrical charges. Then, we move each nodes following the resultant forces. As a result, the distances between the nodes becomes more uniform, the thin rhombus becomes thicker as shown in Fig.3(b), and the properties of the pattern is improved. Another effective technique is disturbing the repetition of similar patterns. As a simple implementation, we slightly move each nodes randomly as shown in Fig.3(c).

To verify the effectiveness of our pattern modification strategy on the original Penrose tiling, we evaluate self-similarity of the proposed patterns, because, for our window matching based approach, it is suitable that all the patterns are unique with large Hamming distance. In addition, since we assume that the projector and camera can be freely installed at scanning process, relative rotational position between the projector and the camera will be changed. Therefore, we check the self similarity for all rotation angles in the evaluation. We use the size 1024x768 of the pattern which is a common resolution on available commercial video projectors, rotate the epipolar line from 0 to 180 degree, and calculate the sum of squared differences (SSD) for disparity range of 3 pixels to 200 pixels using window size 9*9 pixels. Fig.4 shows the lowest SSD value of each angle of epipolar line. Average SSD for a) original Penrose, b) uniform pattern by repulsive force and c) randomness added pattern of 1175, 1413, 1655 as shown in figure as blue dotted lines, respectively. For SSD calculation, since we set the intensity values to be black=0 and white=255, those values infer that at least 4.6 pixels, 5.5 pixels and 6.5 pixels are different from other blocks along the epipolar line.

5 Detailed algorithm of the active stereo

In our method, the reconstruction process consists of mainly three parts. The first step is to calculate the matching cost, considering affine transformation on coarse level. The second is a global optimization, and the third is a disparity calculation for all the pixels at the fine level of two layer pyramid approach.

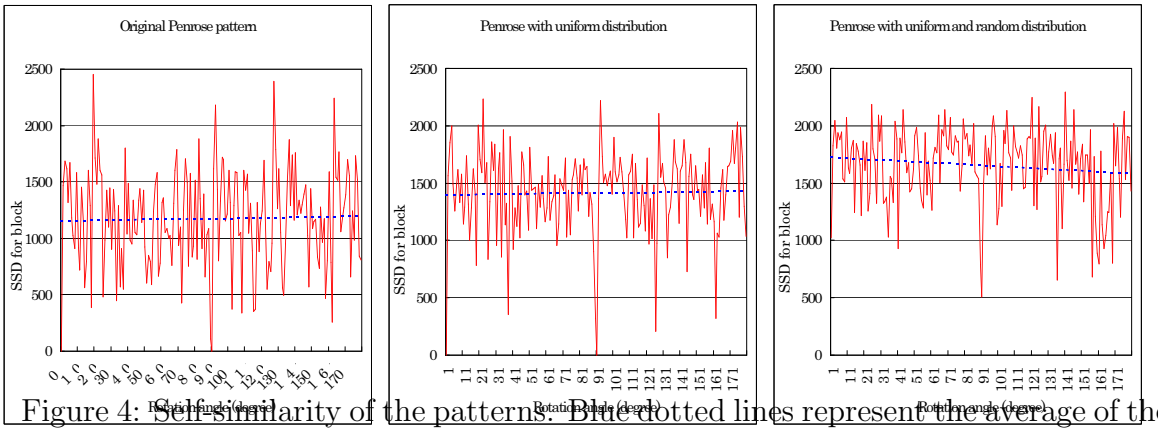


Figure 4: Self-similarity of the patterns. Blue dotted lines represent the average of the values.

As for the input for the reconstruction process, first, we rectify both the captured image and the projected pattern so that the cost calculation process can be conducted along a horizontal line. The disparity search range for the cost calculation is defined by considering the in-focus range of the projector; we assume 10% length of the distance between the projector and the target object in our experiments. After all the matching costs for all the disparities along all the horizontal lines are calculated, final dense 3D shape is calculated by triangulation method.

5.1 Patch based normalization

Since a captured image contains both bright and dark areas, preprocessing is required before cost calculation to retrieve the reliable matching cost. Since the pattern consists of only lines, one solution is to detect lines from the captured image [9, 17, 21]. One drawback of the line detection approach is that the algorithm itself is still under research and it is still an unstable process. Furthermore, a line detection algorithm basically loses some important information for matching costs, *e.g.*, sub-pixel information, gray-level intensity, etc. Therefore, we take another approach to preserve those types of information. In this paper, we apply a window based normalization for local areas. The following linear transformation is conducted for each window of each pixel.

$$I_{new}(x) = (I_{org}(x) - I_{low})\frac{255}{I_{high} - I_{low}}. \tag{1}$$

In the equation, I_{low} and I_{high} represent the lowest and highest value in each window. The window size is defined so as to two times larger than the window size of the matching cost calculation.

5.2 Matching cost calculation with surface orientation estimation

For passive stereo, there are few techniques considering the orientation of the surface of the object [6]. This is natural because high frequency features (higher than window

size) do not exist frequently in the actual scene and color information is sufficient to retrieve good correspondences. Whereas in active stereo, since high frequency patterns are intentionally projected to the object to increase the stability and density of shape reconstruction, the patterns are severely distorted by the surface orientation, which should be resolved. Such matching problems can be solved by using shape invariant feature such as topology information. However, since we use the SSD for cost calculation, another solution is required.

The window size of SSD is usually sufficiently small so that the target surface can be approximated as a patch of plane within the window. Therefore, pattern distortion caused by the surface orientation can be represented by affine transformation with two DOFs. In addition, since we define the matching window size to include at least one node, which has 3 to 5 edges, all the windows are always unique under affine transformations, if the original pattern is unique; this is the case for our pattern as shown in Fig.4.

Since new patterns can be generated by two free parameters, the number of the pattern becomes large, and thus, it is an open problem to efficiently find the correspondences from such large candidates. To overcome this problem, we propose two methods. The first is to precompute the pattern and the second solution is a coarse to fine approach, which is explained in Sec. 5.4. Since the patch size is small and surface orientation is limited in hemisphere, the number of required transformation of the pattern is small in real case. In our experiment, we have found that 42 patterns in the total to represent the affine transformations are sufficient to keep the quality. Therefore, we precompute them before the cost calculation. Fig.5 shows the cost calculation process with precomputed patterns. In the algorithm, the SSD value is calculated by the following equation.

$$SSD(x,d) = \arg\min_{\mathbf{a}} \sum_{x' \in W(x+d)} (I_c(x') - I_p(H_{\mathbf{a}}(x'))^2, \tag{2}$$

where d is a disparity, $W(x)$ is the rectangular patch around x, and $H_a(x')$ is the affine transformation with parameter $\mathbf{a}$. $I_c(\cdot)$ and $I_p(\cdot)$ are the intensities of the camera and projector images, respectively.

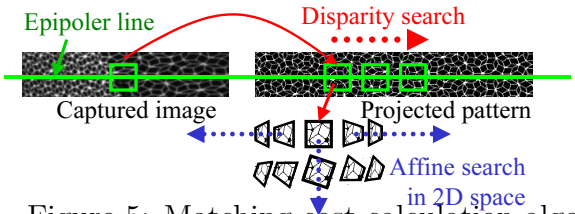


Figure 5: Matching cost calculation algorithm.

5.3 Global optimization

Once all the matching costs are calculated, global optimization is applied to eliminate the small noise which is produced by the wrong correspondences of the patterns. The captured image consists of pixel $p \in V$ and the connections $(p,q) \in U$, where p and q

are adjacent pixels, V is the set of pixels, and U is the set of connections of adjacent pixels. A pixel p has the costs for all the disparities $d_p \in D_p$. We define the energy to find the disparity map as follows:

$$E(D) = \sum_{p \in V} D_p(d_p) + \sum_{(p,q) \in U} W_{pq}(d_p, d_q), \tag{3}$$

where $D = \{d_p | p \in V\}$. $D_p(d_p)$ is the data term of assigning a pixel to disparity d_p. $W_{pq}(d_p, d_q)$ is the regularization term of assigning disparity d_p and d_q to neighboring pixel points. The data term is the SSD calculated by the method described in previous section. The regularization term is defined as follows

$$W_{pq}(d_p, d_q) = |d_p - d_q| \tag{4}$$

The energy is minimized based on belief propagation [5] in this paper.

5.4 Coarse-to-fine technique

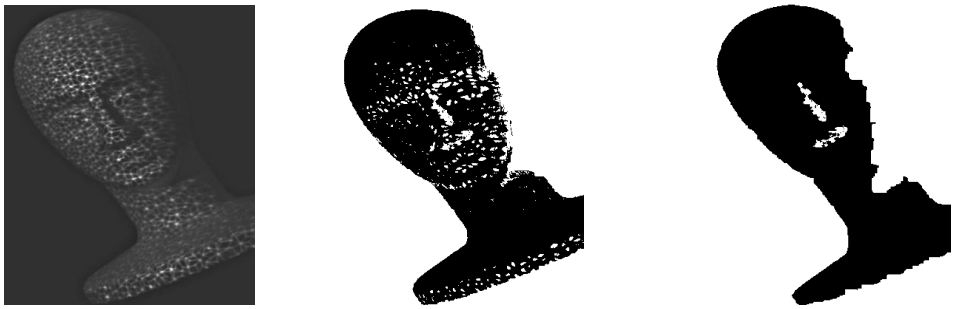
Since the complexity of MRF is NP-hard and complexity of BP is still high and its required memory size is large, we take coarse-to-fine technique (pyramid approach of two layers) to come up with a practical calculation time in our method. For the coarse level, we define the sampling interval as same as the average distance between the nodes of the pattern; the value can be deterministically calculated by the calibration parameters and the projected pattern.

After calculating all the SSDs for coarse level including all the affine transformation, the disparity values are calculated by MRF based global optimization as described in the previous section. At the same time, surface normals for each point of coarse level are estimated.

In the fine level, eight neighboring points of coarse level are checked at the beginning to decide the search range. If the disparity values of all neighboring points are similar and their difference is less than a threshold, just the min and max values are used for the search range. If it is not the case, the area is considered to contain the occluding boundary, and thus, we extend the search range as $(d_{x,y} - r_{thresh}, d_{x,y} + r_{thresh})$ where $d_{x,y}$ is a disparity value on coarse level. Then, disparities for all the pixels are calculated.

5.5 Shape reconstruction near object boundary using edge connectivity

For window matching based approach, reconstructed shapes near object boundary inevitably have noise [10]. In our method, since the pattern of our method has graph representation, we use the edge connectivity to efficiently remove only the noise, but reconstruct the correct shape near the boundary. To use the edge connectivity, explicit approach can be considered, such as detecting edges and reconstructing the shapes only near the edges. With this approach, noises are efficiently removed [9], however, it also decreases the number of reconstructed points because only the few points near the edges are reconstructed. To avoid the problem, we propose the method which implicitly uses the edge connectivity. By using the method, noises are efficiently removed without decreasing the reconstructed points. The actual algorithm is as follows.



(a) Input image (b) Pixel threshold result (c) Only large segments

Figure 6: Calculation process of noise removal near boundary.

1. Check all the pixels on captured image and put label if the intensity is lower than a predefined threshold. Note that, the remaining pixels have certain intensity and considered as a part of the graph (edge or node). Then, connect all the labeled pixels to make independent segments. Example result is shown in Fig.6 (b).
2. Since every edges of the pattern of Penrose tiling are part of the small loops as shown in Fig.3, small segments can be considered as a part of the graph. Therefore, the only segments which are larger than a predefined threshold are considered as a background. Example result is shown in Fig.6 (c).
3. Remove all the 3D points in the segments.

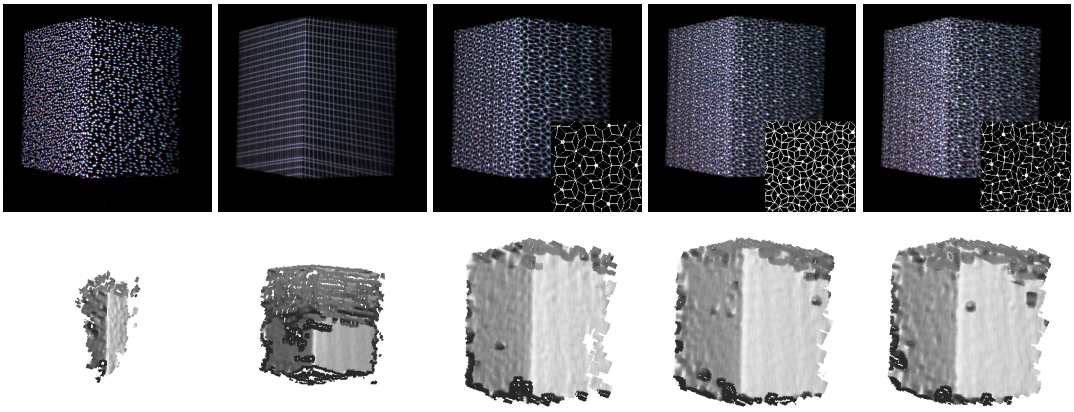
6 Experiment

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.

	(a)random	(b)grid	(c)Penrose	(d)Penrose r1	(e)Penrose r2
RMSE(m)	0.0023	0.0016	0.0018	0.0016	0.0016
Corner angle(deg.)	59.5	58.5	92.0	91.9	91.9
Number of points	10583	31154	26340	28934	28986

Table 1: RMSEs(m) from fitted planes, corner angles, and the number of reconstructed points for results shown in Fig.7.

First, we show the effectiveness of the proposed Penrose tiling pattern by comparing several different patterns. Fig.7 shows the results and Tab.1 shows the RMSEs from the fitted planes, corner angles, and the number of reconstructed points. As shown in Tab.1, case (a) was inaccurate and small in the number of reconstruction. Case (b) was inaccurate in the corner angle (actually, the global position itself was incorrect). In the proposed methods, the number of reconstructed points increased as the randomness was added to the pattern. Therefore, we can conclude that our topology preserving



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

Figure 7: 3D reconstruction results. The top row images are inputs, and the bottom row images are results: (a) the result of a random dot pattern, (b) a grid pattern, (c) a pattern generated from the rhombus tiling(see Fig.3(a)), (d) the rhombus tiling weakly disturbed by noise, and (e) the rhombus tiling strongly disturbed by noise(see Fig.3(c)).

pattern is a promising approach as a single-colored active one-shot scan, and that adding randomness to the pattern can improve performances.

	(d)Phase-shift	(e)Kinect	(g)Grid	(h)Penrose [10]	(i)Our method
RMSE(mm)	-	1.64	1.31	1.11	0.76
Number of points	129,255	3,234	26,849	11,466	40,815

Table 2: RMSE and the number of reconstructed points for the figure shown in Fig.8.

Next, the accuracy of the proposed method was evaluated by capturing the head of the figure as shown in Fig.8. The size of the object was 0.25m high and the distance from the camera was about 1.0m, and the baseline between the camera and the projector for the cases of (c) and (d) is 0.2m. In Fig.8, results by different methods are shown: (d) the temporal-encoding method by projecting phase-shift pattern, (e) Kinect [13], (f) and (g) the spatial-encoding method by projecting single color grid pattern [17], (h) the Penrose pattern [10], and (i) the proposed method. Since the temporal-encoding method (d) has an advantage in terms of its accuracy, we used it as the ground truth for evaluation.

The difference between (f) and (g) is that (f) is the original result with wrong connections between the face and the neck indicated by the red circle, which inevitably occurs on grid pattern without color [17] and the (g) is the result from a manual cut of such wrong connections. Quantitative comparisons are calculated by using the method proposed by Cignoni [4] and shown in Tab.2. From the table, we can confirm that the proposed method (i) gave the best performance on accuracy and our shape reconstruction algorithm for object boundary enabled us to acquire the largest number of reconstructed points among all of the methods.

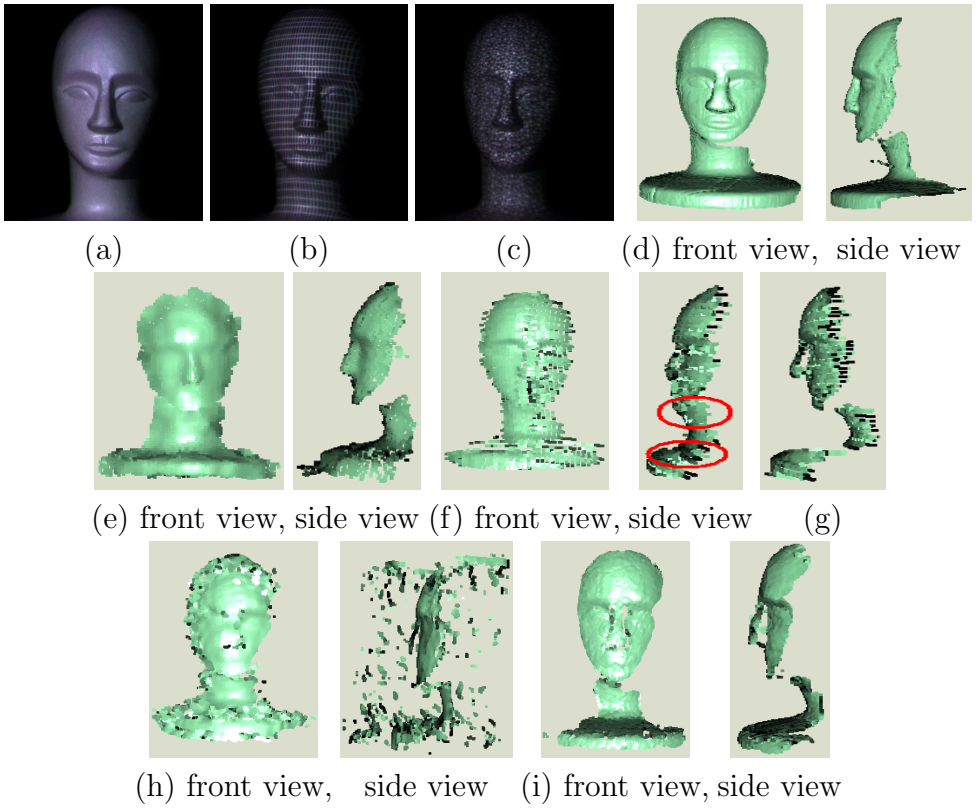


Figure 8: Comparison of other methods: (a) a target object, (b) the object projected by a grid pattern [17], (c) by the proposed pattern, (d) the result of phase-shift as a ground truth, (e) the result of Kinect, (f) and (g) the result from grid pattern, (h) the result from Penrose pattern [10], and (i) our result.

Next, we captured the object which has detailed structure to evaluate the accuracy of the method on small features. For the purpose, we captured a plaster figure as shown in Fig.9(a). The size of the figure was 0.5m height, the distance from the camera was about 1.5m, and the baseline between the camera and the projector is 0.2m. Fig.9(b) is the shape obtained by Gray-code pattern as the ground truth. Fig.9(c) and (d) are the results captured by our method. Fig.9(e) show the error map for all the points of the reconstructed shape. With the map, we can confirm that our method can reconstruct the shape with good accuracy for the entire shape of the object.

Then, we conducted the experiments with more general objects and evaluate the calculation time. Fig.10 shows the results of the captured scenes of hands, faces and a sinusoidal object, respectively. Since the proposed method is a one-shot method, it can generate 3D shapes even if the target object is not static. We can observe some noises near the boundary of the shapes and we consider that this is because surface normals near occluding boundaries are wrongly estimated. Comparing our results (third row) from the results without coarse to fine technique (second row), we can confirm that they have almost the same quality, but our results is smoother. The calculation time is shown in Fig.11. We add the calculation time of the grid based one shot method [17] for comparison. We can clearly see that our coarse-to-fine technique resulted in the

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

Figure 9: 3D reconstruction of a plaster figure: (a) input image, (b) the result of 3D reconstruction by Gray-code pattern, (c) (d) the result by our method and (e) difference between ground truth and our method.

drastic reduction of the computational time, *i.e.*, average 8.5 times faster than ordinary method. However, our method is much slower than the grid based technique, which is a decoding based approach and no search process is required. Since we currently use only CPU and SSD calculation is known to be well adapted to parallel processing, using GPU is promising to speed up and is our important future work.

Finally, since the approach is based on one shot, moving objects can be efficiently captured with our method. Results are shown in Fig.12. With the figure, we can confirm that the motion of the hand, which drastically changes it shapes, is successfully reconstructed by our method.

7 Conclusion

In this paper, efficient and dense 3D reconstruction method from a single image using single-colored static pattern is proposed. The method utilizes topology information to achieve wider base-line with stable reconstruction compared to the previous methods. We also propose an efficient stereo algorithm to increase the stability by solving the affine transformation of the pattern by estimating the surface normal of the object. To shorten the time, a coarse-to-fine strategy is implemented. At the coarse level, MRF based optimization is applied by BP to improve the result. In the experiments, we evaluated the accuracy of our method compared to the state-of-the-art one-shot scan techniques and prove the strength of our method. We also compare the calculation time to confirm the effectiveness of our coarse-to-fine strategy. In the future, further research using GPU implementation for faster calculation is planned.

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).

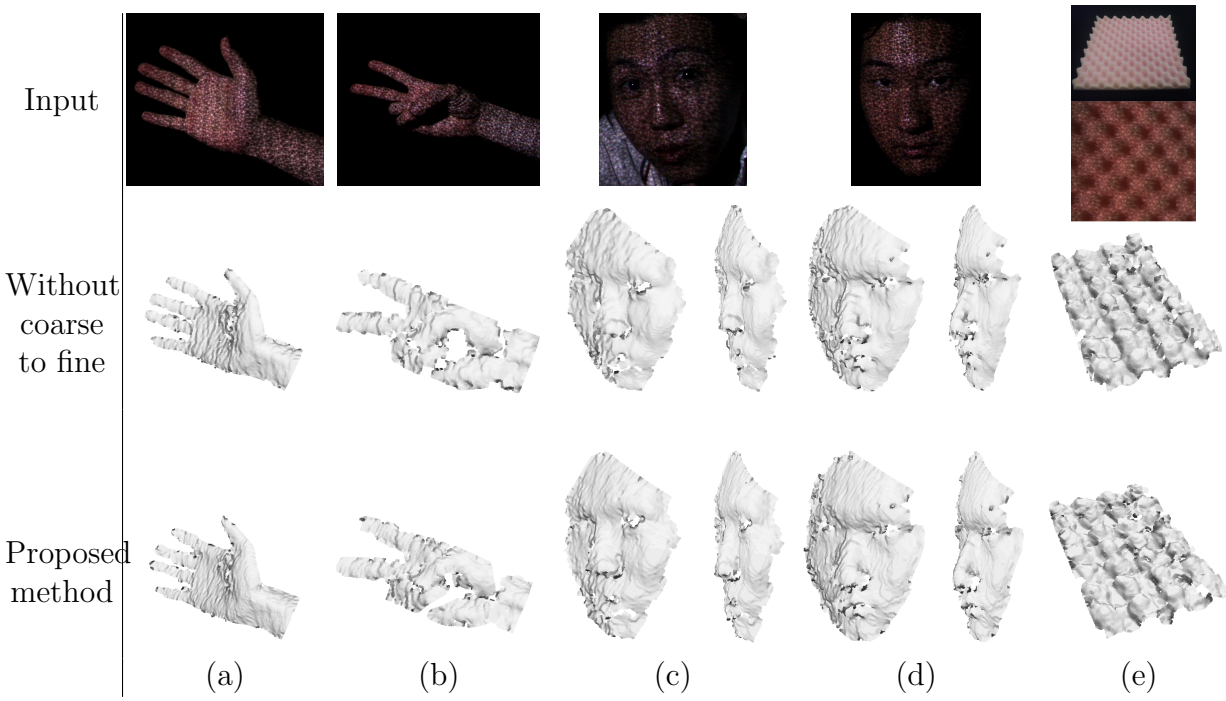


Figure 10: 3D reconstruction results of general objects. The top row images are inputs, and the bottom row images are the results: (a) an open hand, (b) a scissor hand, (c) and (d) faces, and (e) a sinusoidal object.

References

[1] 3dMD, "3dMDface System", URL `http://www.3dmd.com`.

[2] Artec, "United States Patent Application 2009005924", (2007j).

[3] Canesta, Inc., "CanestaVision EP Development Kit", (2010), http://www.canesta.com/devkit.htm.

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

[5] P. Felzenszwalb, D. Huttenlocher, "Efficient belief propagation for early vision", *IJCV*, **70**, pp. 41–54 (2006).

[6] Y. Furukawa, J. Ponce, "Accurate, dense, and robust multi-view stereopsis", *in CVPR* (2007).

[7] M. Gardner, *Penrose Tiles to Trapdoor Ciphers*, Cambridge University Press (1997).

[8] C. Je, S. W. Lee, R.-H. Park, "High-contrast color-stripe pattern for rapid structured-light range imaging", *in ECCV*, volume 1, pp. 95–107 (2004).

[9] H. Kawasaki, R. Furukawa, , R. Sagawa, Y. Yagi, "Dynamic scene shape reconstruction using a single structured light pattern", *in CVPR*, pp. 1–8 (June 23-28 2008).

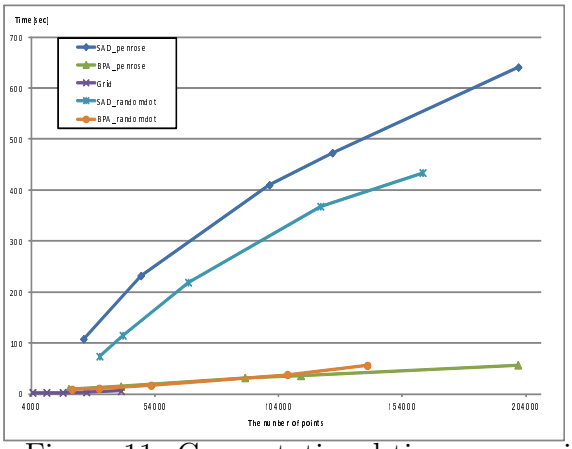


Figure 11: Computational time comparison

Frame 1 Frame 3 Frame 5 Frame 7 Frame 9

Figure 12: Sequence of 3D reconstruction results of a moving object (palm of a hand).

[10] H. Kawasaki, H. Masuyama, R. Sagawa, R. Furukawa, "Single color one-shot scan using topology information", *ECCV2012 Workshop*, **LNCS7585(PartIII)**, pp. 486–495 (10 2012).

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

[12] M. Maruyama, S. Abe, "Range sensing by projecting multiple slits with random cuts", *in SPIE Optics, Illumination, and Image Sensing for Machine Vision IV*, volume 1194, pp. 216–224 (1989).

[13] Microsoft, "Xbox 360 Kinect", (2010), http://www.xbox.com/en-US/kinect.

[14] S. G. Narasimhan, S. J. Koppal, , S. Yamazaki, "Temporal dithering of illumination for fast active vision", *in Proc. European Conference on Computer Vision*, pp. 830–844 (October 2008).

[15] S. Rusinkiewicz, O. Hall-Holt, M. Levoy, "Real-time 3D model acquisition", *in Proc. SIGGRAPH*, pp. 438–446 (2002).

[16] R. Sagawa, H. Kawasaki, R. Furukawa, S. Kiyota, "Dense one-shot 3D reconstruction by detecting continuous regions with parallel line projection", *in ICCV*, pp. 1911–1918 (2011).

[17] R. Sagawa, Y. Ota, Y. Yagi, R. Furukawa, N. Asada, H. Kawasaki, "Dense 3d reconstruction method using a single pattern for fast moving object", *in ICCV* (2009).

[18] J. Salvi, J. Batlle, E. M. Mouaddib, "A robust-coded pattern projection for dynamic 3D scene measurement", *Pattern Recognition*, **19(11)**, pp. 1055–1065 (1998).

[19] J. Salvi, J. Pages, J. Batlle, "Pattern codification strategies in structured light systems", *Pattern Recognition*, **37(4)**, pp. 827–849 (4 2004).

[20] J. Tajima, M. Iwakawa, "3-D data acquisition by rainbow range finder", *in ICPR*, pp. 309–313 (1990).

[21] A. O. Ulusoy, F. Calakli, G. Taubin, "One-shot scanning using de bruijn spaced grids", *in The 7th IEEE Conf. 3DIM*, pp. 1786–1792 (2009).

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

[23] T. Weise, B. Leibe, L. V. Gool, "Fast 3D scanning with automatic motion compensation", *in CVPR* (2007).

[24] L. Zhang, B. Curless, S. Seitz, "Rapid shape acquisition using color structured light and multi-pass dynamic programming", *in 3DPVT*, pp. 24–36 (2002).

[25] S. Zhang, P. Huang, "High-resolution, real-time 3D shape acquisition", *in Proc. Conference on Computer Vision and Pattern Recognition Workshop*, p. 28 (2004).