Single Color One-Shot Scan Using Topology Information

Hiroshi Kawasaki[1], Hitoshi Masuyama[1], Ryusuke Sagawa[2], and Ryo Furukawa[3]

[1] Kagoshima University, Kagoshima, Japan
[2] AIST, Tsukuba, Japan
[3] Hiroshima City University, Hiroshima, Japan

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 a moving object. Since a 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 used for reconstruction, positional information of a pixel on 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. In this paper, we propose a technique to solve the problem by using topological information instead of colors. Our technique successfully realizes one-shot scan with monochrome pattern.

1 Introduction

Importance of shape capture of moving objects is rapidly increasing. For example, recently, inexpensive scanning devices developed for entertainment purposes made a great success to achieve the device-free interface [1]. Because their purpose of the scanner is mainly on a motion capture, their accuracy and density are relatively low, compared to existing range sensors for industrial purposes. If high accuracy with dense resolution is realized on such scanners, they become more useful for various purposes, *e.g.*, medical application and fluid analysis.

There are several methods exist for capturing moving objects with active scanning techniques, such as stereo based methods or time-of-flight (TOF) methods. Especially, structured-light stereo methods are suitable for capturing moving objects and have been widely researched [1–4]. Structured-light methods are usually categorized into two types: temporal-encoding and spatial-encoding methods. Since the spatial-encoding method just requires a single input for reconstruction (*a.k.a.* one-shot scan), it is ideal for capturing moving objects. Therefore, many researches have been conducted with spatial-encoding methods [5]. However, since they require certain areas to encode the position of the pixel of the projected pattern, the resolution tends to be low and reconstruction becomes unstable with an inevitable turbulence of color. One efficient way to encode information on the surface of the object is to use an epipolar geometry. By using

A. Fusiello et al. (Eds.): ECCV 2012 Ws/Demos, Part III, LNCS 7585, pp. 486–495, 2012.

it, ambiguity can be decreased from 2D to 1D and stability can be improved drastically. To use the epipolar constraint, stripe pattern which is perpendicular to the epipolar line is commonly used [6–8]. Grid pattern is also used to increase the stability [9, 3, 10]. Although such approaches greatly ease the problem, since color information is easily affected by texture, material and lighting conditions, results are still unstable in a real scene.

To avoid color problems, several methods are proposed for efficient spatial encoding without using colors, such as dot patterns or grid patterns [1]. Even though, there still remain several problems, *i.e.*, inaccuracy by short base line and sparse reconstruction. In this paper, we propose a one-shot scanning method which can solve the aforementioned problems with the following approaches.

Topology Information for Wide Base-Line Stereo: To increase the stability on retrieving correspondences, we propose a topology information instead of color information. We use a graph representation for the pattern and 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 the topology information is preserved during geometric transformation, robust correspondences with wide base-line can be realized.

Geometric Information for Dense Reconstruction: Topology information can only be applied sparsely on the pattern, we use geometric information to increase the density of reconstruction. As for the implementation, a small window is used to calculate matching scores using geometric information for each pixel. Unlike the topology information, the matching score is sensitive to the geometric transformation, we also estimate the surface normal for each pixel. Although such pixels are reconstructed unstably, a global optimization technique is conducted.

Global Optimization to Decrease Wrong Reconstruction: Since the technique is based on stereo, MRF based global optimization technique can be applied. In our method, the matching scores of each pixel especially with geometric information tend to have several local minima; such multiple candidates are efficiently solved by global optimization. We use belief propagation method in the paper.

2 Related Work

Triangulation based methods (*e.g.*, light-sectioning method or stereo method) and time-of-flight(TOF) based methods are widely known for active measurement. Since many TOF based systems use point lasers, 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[11]. However, the present systems are easily disturbed by other light sources and the resolution is low.

With regard to triangulation based methods, many methods 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, such as video projector, is a simple solution to reduce the time to scan. However, unlike

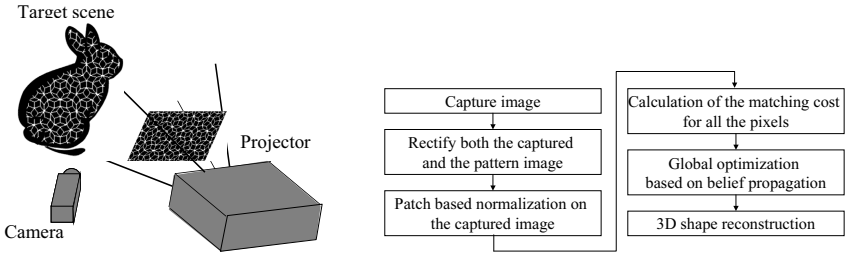


Fig. 1. (a) System configuration where a graph-based pattern is projected from the projector and captured by the camera, and (b) an algorithm overview

a point or a line light sources, there is an ambiguity on correspondences. For solution, typically two methods are known, *i.e.*, temporal-encoding or spatial-encoding methods[6].

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 the problem; by capturing with high frequencies [12–14]. Although it is reported that some works can capture around 100 FPS by combining motion compensation, since these methods require multiple frames, the quality of the results is degraded if object moves fast.

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 [15, 16], dotted lines with unique modulations of dots [17, 18], 2D area information for encoding [19, 1], or connections of grid patterns [20, 2–4]. However, no method has achieved a sufficient performance in all aspects of precision, resolution, and stability.

In the paper, we propose a simple technique to solve the aforementioned problems using the new pattern which uses topology information. With our technique, all the problems are not solved, however, some promising aspects can be shown for future direction of one-shot scan with single color.

3 Overview and System Settings

Our system consists of a single projector and a camera. The projector casts a static pattern as shown in Fig.1(a). The pattern consists of lines (edges) and intersections (nodes) which make a graph representation (details are described in Sec.4). Since the pattern is static, no synchronization is required.

Overview of our algorithm is shown in Fig.1(b). First, we rectify both the captured image and the projected pattern. Then, we normalize the captured

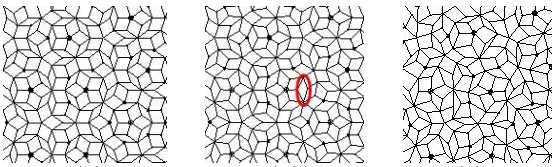


Fig. 2. Graph pattern generated by Penrose tiling: (a) a basic pattern, (b) a basic pattern that is modified so that thin rhombuses (one of them is marked by a red ellipse) are enlarged, and (c) a pattern in which the positions of the nodes are disturbed to reduce the repetition of similar patterns

image for better calculation of the matching cost. In this process, since environmental lighting condition and texture is not uniform, adaptive normalization is applied. In the next step, matching costs are calculated for each pixel. We also estimate the surface normal in this step. Using the cost, global optimization is conducted using BP. Finally, the depths for all the pixels are reconstructed using the estimated disparity for each pixel.

4 Topology Preserving Pattern

In one-shot active stereo, the pattern projected to the target is important 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 features, thus, it is not appropriate for a pattern for stereo matching as it is. One of the possible solution for this is to use a pattern for 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 the present work, we propose to use a pattern generated by Penrose tiling[21], which has plenty of such features.

Penrose tiling is a kind of tiling (filling a plane with some geometric shapes without overlaps nor gaps) that can be generated by a small number of tiles. The generated patten is known to have no translational symmetry. Among several kinds of Penrose tiling, we use a pattern that are generated by two kinds of rhombuses[21] (rhombus tiling). The generation can be easily achieved using a recursive algorithm. An example of rhombus tiling is shown in Fig.2(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 an 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. About the proposed pattern, the orders of nodes are 3, 4, or 5. This means that, although this graph has several orders of nodes, the maximum order of a node is limited to be 5. Moreover, the density of nodes in the pattern is uniform. This is useful to achieve dense reconstruction at actual measurement. From the abovementioned properties, the graph generated by rhombus tiling can be considered to have good properties as an active stereo pattern.

The minimum angle of corners of the rhombuses, which is shown in Fig.2(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 some homographic deformations. 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.2(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.2(c).

5 Reconstruction by Stereo with Regularization

In our method, reconstruction process consists of mainly two parts. The first one is a matching cost calculation part and the other is a global optimization part.

As the input for 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. Then, disparity search range for the cost calculation is defined by considering the in-focus range of the projector; it is usually 10% length of the distance between the projector and the target. After all the matching costs for all the disparities along all the horizontal lines are calculated, global optimization is applied to get the final 3D shape.

5.1 Matching Cost Calculation

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 may consider that the line detection can be a solution. It is true and that can be used as the normalization process, however, the algorithm itself is still under research. Furthermore, a line detection algorithm basically loses some important information for matching cost, such as sub-pixel information with gray-scale intensity. Therefore, we take another approach to preserve those information. In this paper, we apply window based normalization for local areas. The following linear transformation is conducted for each window defined for each target pixel.

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

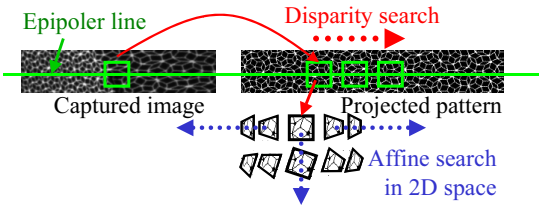


Fig. 3. Matching cost calculation algorithm

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 for the matching cost calculation.

Matching Cost Calculation with Surface Orientation Estimation. For passive stereo, there are a few techniques considering the orientation of the surface of the object [22]. 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 the captured image, the patterns are severely distorted by the surface orientation, which should be resolved.

In this paper, we calculate the matching cost using SSD with a small window and the size of the window is defined so that at least one intersection point is included. In such case, the window size is relatively small compared to the image size, and thus, pattern distortion caused by the orientation of the surface can be represented by affine transformation with two DOFs. Since each window contains at least three lines which share the intersection point, pattern is always unique under affine transformations. Therefore, we can estimate the surface orientation for each point independently. In our method, instead of applying optimization technique to find the solution, we conduct full search to estimate the parameter. There are mainly two reasons for this. First, since topology pattern has several local minima, simple optimization techniques, such as the least descent method, sometimes fail. Secondly, because the patch size is small, the number of variety of pattern transformation could be small. Therefore, for actual implementation, we precompute the pattern with a limited number of affine transformation and find the best match to estimate the surface normal. In our experiment, we find that total 42 patterns (6 and 7 for each parameter) are sufficient to produce enough quality. Fig.3 show the cost calculation process and 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.

5.2 Global Optimization

Once all the matching costs are calculated, global optimization is applied to eliminate the small noise which is produced by the self-similarity 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

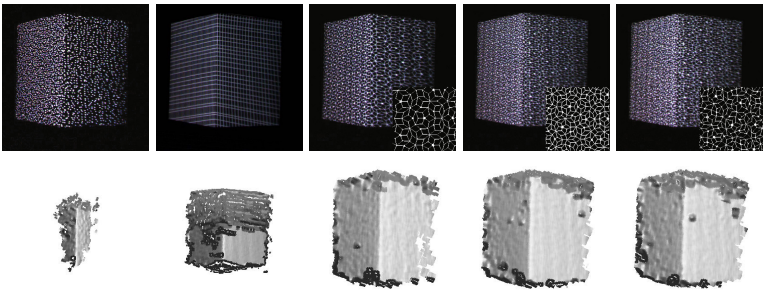
$$W_{pq}(d_p, d_q) = |d_p - d_q| \tag{4}$$

The energy is minimized based on belief propagation [23] in this paper.

6 Experiment

We applied a camera of 1600×1200 pixels, a projector of 1024×768 pixels and a PC with Intel Core i7 2.93GHz/NVIDIAGeForce 580GTX.

First, we show the effectiveness of the proposed topology pattern by comparing with several different patterns. Fig.4 shows the results and table 1 shows the RMSEs from the fitted planes, corner angles, and the number of reconstructed points. As shown in table 1, case (a) was inaccurate and small in the number



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

Fig. 4. 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.2(a)), (d) the rhombus tiling weakly disturbed by noise, and (e) the rhombus tiling strongly disturbed by noise(see Fig.2(c)).

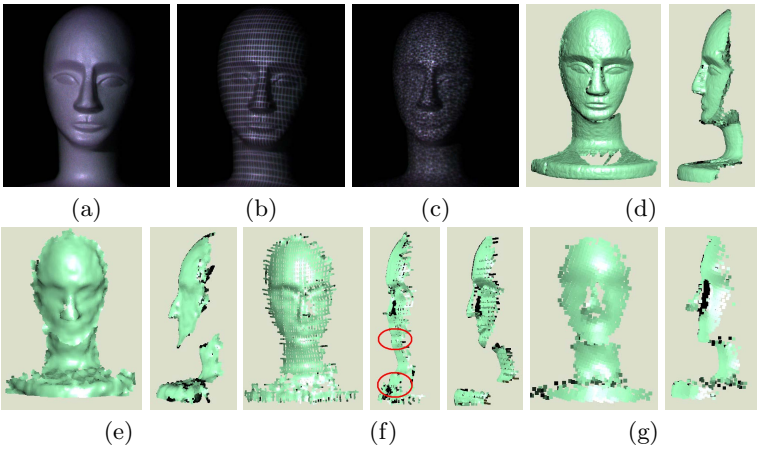


Fig. 5. Comparison of other methods: (a) a target object, (b) the object projected by a grid pattern [3], (c) by the proposed pattern, (d) the result of phase-shift as a ground truth, (e) the result of Kinect, (f) the result from (b), and (g) the result from (c)

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

Table 1. RMSEs(m) from fitted planes, corner angles, and the number of reconstructed points for results shown in Fig.4

	(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

Next, the accuracy of the proposed method was evaluated by capturing the head of the figure as shown in Fig.5. The size of the object was 0.25m high and the distance from the camera was about 0.6m. In Fig.5, results by different methods are shown: (d) the temporal-encoding method by projecting phase-shift pattern, (e) Kinect, (f) the spatial-encoding method by projecting single color grid pattern [3], and (g) the proposed method. Since the temporal-encoding method (d) has an advantage in terms of accuracy, we used it as the ground truth for evaluation. In figure (f), we put two results; the left one is the result with wrong connections between the face and the neck indicated by the red circle, which inevitably occurs on grid pattern [3] and the right where such wrong connections are cut. The differences between points are calculated by using a

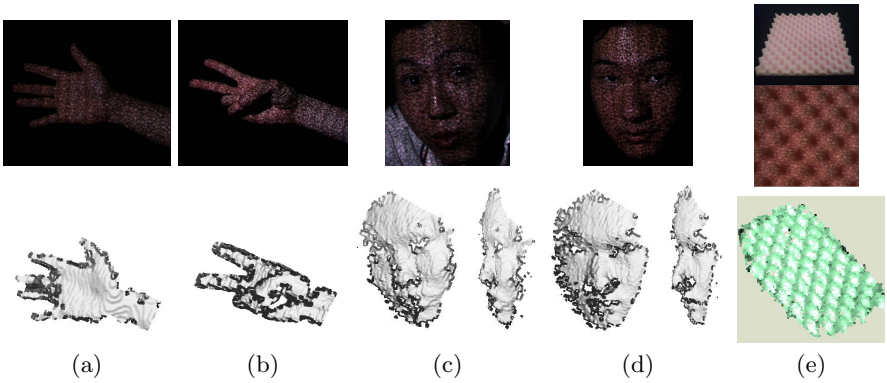


Fig. 6. 3D reconstruction results of general objects. The top row images are inputs, and the bottom row images are the results: (a) an open hand, (b) a scissor hand, (c) and (d) faces, and (e) a sinusoidal object.

method proposed by Cignoni [24]. The RMSEs from ground truth are (e) 0.4mm, (f) left 2.3mm, (f) right 0.09mm, and (g) 0.07mm, respectively. We can confirm that the propose method (g) gave the best performance. However, the number of the reconstructed points is smaller than other methods. We consider that this is mainly because our method does not use the information of the connection of the pattern. This is our future work to solve the problem. Calculation time of our method was around 1min. to 5min. per image. Speeding up the calculation time is also our important future work.

Finally, we show the results of more general objects. Fig.6 show the results of the captured scenes of hands, faces and a sinusoidal object, respectively. Since the proposed method is one-shot method, it can generate 3D shapes even if the target object is not static. Here, you can see some noises near the boundary of the shapes. We consider that this is because surface normals near occluding boundaries are wrongly estimated.

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 a geometric information to increase the density by solving the affine transformation of the pattern. At the final reconstruction step, BP technique is used to integrate both topology information and geometry based techniques. In the experiments, we evaluated the accuracy of our method compared to the state-of-the-art one-shot scan techniques and proves the strength of our method. Several directions of the future research are presented.

References

1. Microsoft: Xbox 360 Kinect (2010), http://www.xbox.com/en-US/kinect
2. Kawasaki, H., Furukawa, R., Sagawa, R., Yagi, Y.: Dynamic scene shape reconstruction using a single structured light pattern. In: CVPR, pp. 1–8 (2008)
3. Sagawa, R., Ota, Y., Yagi, Y., Furukawa, R., Asada, N., Kawasaki, H.: Dense 3d reconstruction method using a single pattern for fast moving object. In: ICCV (2009)
4. Ulusoy, A.O., Calakli, F., Taubin, G.: One-shot scanning using de bruijn spaced grids. In: The 7th IEEE Conf. 3DIM (2009)
5. Salvi, J., Pages, J., Batlle, J.: Pattern codification strategies in structured light systems. Pattern Recognition 37, 827–849 (2004)
6. Salvi, J., Batlle, J., Mouaddib, E.M.: A robust-coded pattern projection for dynamic 3D scene measurement. Pattern Recognition 19, 1055–1065 (1998)
7. Je, C., Lee, S.-W., Park, R.-H.: High-Contrast Color-Stripe Pattern for Rapid Structured-Light Range Imaging. In: Pajdla, T., Matas, J. (eds.) ECCV 2004, Part I. LNCS, vol. 3021, pp. 95–107. Springer, Heidelberg (2004)
8. Zhang, L., Curless, B., Seitz, S.: Rapid shape acquisition using color structured light and multi-pass dynamic programming. In: 3DPVT, pp. 24–36 (2002)
9. Furukawa, Y., Ponce, J.: Dense 3D motion capture from synchronized video streams. In: CVPR (2008)
10. Sagawa, R., Kawasaki, H., Furukawa, R., Kiyota, S.: Dense one-shot 3D reconstruction by detecting continuous regions with parallel line projection. In: ICCV (2011)
11. Canesta, Inc.: CanestaVision EP Development Kit (2010), http://www.canesta.com/devkit.html
12. Rusinkiewicz, S., Hall-Holt, O., Levoy, M.: Real-time 3D model acquisition. In: Proc. SIGGRAPH, pp. 438–446 (2002)
13. Weise, T., Leibe, B., Van Gool, L.: Fast 3D scanning with automatic motion compensation. In: CVPR (2007)
14. Narasimhan, S.G., Koppal, S.J., Yamazaki, S.: Temporal Dithering of Illumination for Fast Active Vision. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part IV. LNCS, vol. 5305, pp. 830–844. Springer, Heidelberg (2008)
15. Tajima, J., Iwakawa, M.: 3-D data acquisition by rainbow range finder. In: ICPR, pp. 309–313 (1990)
16. Zhang, S., Huang, P.: High-resolution, real-time 3D shape acquisition. In: Proc. Conference on Computer Vision and Pattern Recognition Workshop, p. 28 (2004)
17. Maruyama, M., Abe, S.: Range sensing by projecting multiple slits with random cuts. In: SPIE Optics, Illumination, and Image Sensing for Machine Vision IV, vol. 1194, pp. 216–224 (1989)
18. Artec: United States Patent Application 2009005924 (2007j)
19. Vuylsteke, P., Oosterlinck, A.: Range image acquisition with a single binary-encoded light pattern. IEEE Trans. on PAMI 12, 148–164 (1990)
20. Koninckx, T., Van Gool, L.: Real-time range acquisition by adaptive structured light. IEEE Transaction Pattern Analysis Machine Intelligence 28, 432–445 (2006)
21. Gardner, M.: Penrose Tiles to Trapdoor Ciphers. Cambridge University Press (1997)
22. Furukawa, Y., Ponce, J.: Accurate, dense, and robust multi-view stereopsis. In: CVPR (2007)
23. Felzenszwalb, P., Huttenlocher, D.: Efficient belief propagation for early vision. IJCV 70, 41–54 (2006)
24. Cignoni, P., Rocchini, C., Scopigno, R.: Metro: measuring error on simplified surfaces. Computer Graphics Forum 17, 167–174 (1998)