Robust and Accurate One-shot 3D Reconstruction by 2C1P System with Wave Grid Pattern

Nozomu Kasuya
Kagoshima University
Kagoshima, Japan
AIST
Tsukuba, Japan
nozomu.kasuya@aist.go.jp

Ryusuke Sagawa
AIST
Tsukuba, Japan
ryusuke.sagawa@aist.go.jp

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

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

Abstract

In this paper, we propose an active 3D reconstruction method with two cameras and one projector (2C1P) system for capturing moving objects. The system reconstructs the shapes from a single frame of each camera by finding the correspondence between the cameras and the projector based on projecting wave grid pattern. The projected pattern gives the constraint of correspondence between the two cameras in addition to between a projector and a camera. The proposed method finds correspondence by energy minimization on graphs constructed by detecting a grid pattern in camera images. Since the graphs of two cameras are connected as a single graph by using the constraint between cameras, the proposed method simultaneously finds the correspondences for two cameras, which contributes to the robustness of correspondence search. By merging range images created by the correspondence of each camera, we reduce the occluded area compared to the case of one camera. Finally, the proposed method optimizes the shape as three-view stereo to improve the accuracy of shape measurements. In the experiment, we show the effectiveness of using two cameras by making comparison with the case of one camera.

1. Introduction

Recently, capturing the shapes of objects in motion has become popular in various applications. For example, gaming device that can capture a human motion in real-time and create a device-free interface made more than ten million sales [8]. Another example is a performance capture [2] which acquires surface of the shape and kinematic information for fields such as 3D movies, motion analysis, apparel manufacturing.

Various approaches including multiview stereo, TOF camera and its variations devices have been proposed to obtain surface shapes of moving objects. Among them, structured-light stereo systems are suitable for capturing moving objects and have been widely researched [15, 5, 19, 11]. To capture moving objects with high FPS, spatial-encoding methods that use a single input image for reconstruction (*a.k.a.* one-shot scan) have been proposed. Since structured-light systems are active methods, they have several advantages. For instance, simplification of correspondence search, robustness for capturing texture-less objects, and simple setup with one camera and one projector.

One of the disadvantages of the spatial-encoding methods is that the ambiguity on correspondence. Since these methods find the correspondence between camera and projector from a single image and the information embedded in the single projection pattern is limited, accuracy, density, or robustness of measurements will be sacrificed in the spatial-encoding methods. The goal of this paper is to improve the robustness of correspondence by adding a camera that gives additional constraint.

Additionally, surfaces that are measurable by the methods based on triangulation is restricted to the ones visible from more than two cameras or projectors. In the cases of structured-light systems, even if the pattern is projected on the surface, the surface parts occluded from the camera cannot be measured. Using two cameras contributes to the

efficient usage of the projected pattern by reducing the occluded area.

In this paper, we extend one of the spatial-encoding methods proposed by Sagawa *et al*. [13], which reconstructs shapes by projecting a static grid pattern of wave lines. In their previous approach [4], they proposed a setup with one camera and two projectors (1C2P). The method casts parallel lines from each projector and makes the grid pattern on object's surface. Since the grid is necessary for reconstruction, the part only visible from one projector must be connected to the area that have grid pattern. In contrast, we use two cameras and one projector. Our approach has the following advantages:

1. Improve the robustness of finding correspondences by using camera-camera information in addition to camera-projector information.

2. Improve the accuracy of correspondences by three-view geometry of the 2C1P system.

3. Reduce the occluded area and increase the efficiency of the projected pattern.

The method proposed in [13] finds the correspondence as the energy minimization problem of a graph which is constructed by detecting the grid pattern in the camera image. Since the detected graph should satisfy the geometrical constraint as the projection of the grid pattern, the constraint is used to find the correspondence. If we have two cameras and one projector, the additional geometrical constraint between them exists, which is known as trilinear constraint in multiple view geometry. We introduce this constraint to the energy minimization of the graph in this paper.

2. Related Work

Structured-light systems are categorized into two types, temporal-encoding or spatial-encoding method [14]. In a temporal-encoding method, multiple patterns of illuminations are projected, and the corresponding 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 [10], reducing the required number of patterns by using phase patterns[21], or using DMD patterns[9]. Another approach is a space-time stereo, where multiple cameras are used with temporally varying the illumination [22]. The method in [21] used two cameras to remove phase ambiguity, which is regarded as the combination of passive stereo and structured-light system, while the pattern are not designed to give unique correspondence in space-time stereo. Although it is reported that some works can capture around 100 FPS by combining motion estimation, since these methods require multiple frames, the quality of the results degrade if the object moves fast.

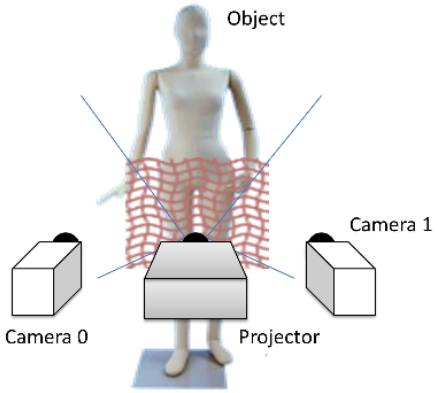


Figure 1. The proposed scanning system uses two cameras and one projector (2C1P).

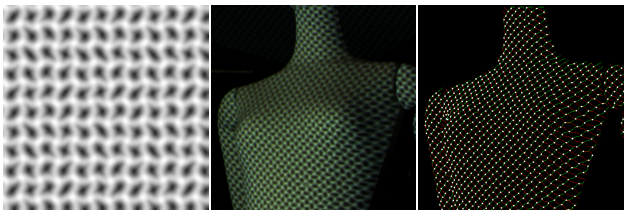
A spatial-encoding method uses a static pattern and usually requires just a single image, and thus, it is suitable to capture dynamic scenes. Many methods have been proposed to solve the problems; *e.g*., using multiple lines with globally-unique color combinations [18, 23], dotted lines with unique modulations of dots [7, 1], 2D area information for encoding [20, 8], using the phase of a fringe pattern [17, 16], or connections of grid patterns [6, 5, 12, 19]. However, no method has achieved sufficient performances in all aspects of precision, resolution, and stability. While projecting wave grid pattern [13] is a solution to the problem, it has restriction on the relative position between projector and camera. In this paper, we propose a method to use two cameras for finding correct correspondence even in the case that the cameras are at an inappropriate position.

3. Wave Grid-based Active Stereo with Two Cameras

The proposed method uses two cameras and one projector (2C1P) as shown in Fig.1. To reconstruct the shapes of moving objects, we extend the method proposed in [13], the system which consists of one camera and one projector (1C1P). We first explain the basic method and then describe our new approach that introduces camera-camera constraint.

3.1. 1C1P System with Wave Grid Pattern

The basic system [13] has a single projector and a camera. Since they are assumed to be calibrated, the intrinsic and extrinsic parameters of the devices are known. The projector casts a static pattern as shown in Fig.2 (a). The pattern consists of vertical and horizontal sinusoidal curves to create grid shape. Since the pattern is static with single



(a) (b) (c)

Figure 2. (a) is an example of a wave grid pattern. (b) is an input image. (c) is the grid graph generated by line detection.

color, no synchronization is required and high FPS scanning is possible.

First, we detect a grid pattern from captured images by the method based on belief propagation (BP) proposed in [12], which extracts vertical and horizontal curves in the camera image separately by discriminating their direction. Next, the grid graph is constructed from the detected pattern. The nodes of the graph are calculated as the intersection points of vertical and horizontal curves, and the edges are given by connecting the neighboring intersection points of a detected curve as shown in Fig.2(c). Then, the epipolar line on the projected pattern is calculated to find the candidates of correspondence to be used for each node. The best correspondences are chosen from the candidates by energy minimization based on BP. The depths for all pixels of the camera are interpolated by using the correspondence of the graph nodes. Finally, the depths are optimized by matching the points between the projected pattern and the captured image, and the dense 3D shapes are reconstructed from the depths.

The method based on projecting wave grid pattern uses the constraint that the nodes connected by the edges should be on the same line in the projector pattern, which is called coplanar constraint in [5]. In the case that the epipolar line in the projector image is parallel to the direction of wave lines, the wave lines do not curve regardless of the shape of the object, which indicates that the connectivity gives no constraint. While a 1C1P system cannot solve the problem in such cases, the proposed method can find the correct correspondence by using the additional constraint given by a 2C1P system. Not even in the extreme case described above, the camera-camera constraint is effective for robust correspondence search.

3.2. Adding Geometrical Constraint between Cameras

The proposed method uses the constraint placed between the cameras as additional information to find correspondences. As described in the following section, the method determines the correspondences based on the energy minimization on the grid graph. Therefore, we introduce the additional constraints as the edges that connect graphs of two cameras.

Fig.3 shows how to generate edges between two graphs. First, we detect the wave line on the camera images and create the grid graphs. Next, let us consider to determine the corresponding point in the projected pattern of a node p_0 of camera 0, which is a grid point where two lines intersect. The candidates of the corresponding points $t_{p_0} \in T_{p_0}$ are the intersection points of the pattern on the epipolar line of p_0 in the projector image, where T_{p_0} are the set of the candidates for the node p_0. If we assume the correspondence of p_0 and t_{p_0}, the 3D coordinates $P_{3D}(t_{p_0})$ for the nodes p_0 are calculated by triangulation between camera 0 and the projector. Next, the projection of the 3D points $P_{3D}(t_{p_0})$ onto the image of camera 1 is $P_1(t_{p_0})$ as shown in Fig.3. If the node p_1 of camera 1 satisfies the following two conditions, p_0 and p_1 can be the corresponding points.

$$D(p_1, P_1(t_{p_0})) < \theta \quad \text{and} \quad t_{p_0} \in T_{p_1}, \tag{1}$$

where $D(a, b)$ is the distance between points a and b, θ is the radius of the search area near the node from $P_1(t_{p_0})$, and T_{p_1} is the set of candidates of t_{p_1}. Four $P_{3D}(t_{p_0})$ are projected onto camera 1 in Fig.3. Since the leftmost $P_1(t_{p_0})$ has no nodes in the search area, no candidate of correspondence is found. While the rightmost one has a node p_1 in the search area, the node does not have the same candidate t_{p_0} in T_{p_1}. Since the middle two projections satisfy the above condition, their nodes are connected to p_0. Once the edges between two cameras connect their graphs, they become a single graph, which enables to simultaneously optimize the correspondences search of two cameras.

In this method, some incorrect edges can be generated. The second projection from the left Fig.3 is an example of incorrect edges, which are not on the object's surface. However, if the node has both correct and incorrect edges, the total cost of BP is not affected by the incorrect edge (details are described in Sec.3.3). In the case that a node has only incorrect edges, the candidate of correspondence can be determined as false in the process of BP if the number of incorrect edges is not sufficiently small.

We find the candidates of correspondence from camera 1 to camera 0 in the same manner, too.

3.3. Simultaneous Correspondence Search of Two Cameras as Energy Minimization of a Single Graph

Now, we have obtained a single grid graph for two cameras by detecting lines and reprojecting points to another camera. The next step is to find the best combinations of correspondences, which is determined by the energy minimization on the grid graph.

The grid graph consists of nodes $p_0 \in V_0$,$p_1 \in V_1$, edges by line detection $(p_0, q_0) \in U_0, (p_1, q_1) \in U_1$, and edges

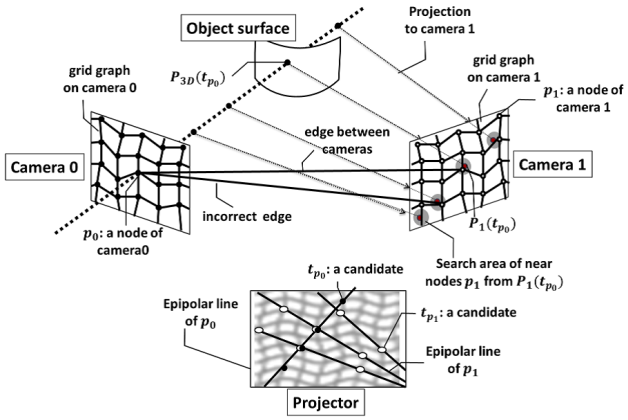


Figure 3. Grid points $t_{p_0} \in T_{p_0}$ of the projector pattern are the candidate of correspondence of a node p_0 of camera 0. $P_{3D}(t_{p_0})$ is calculated by using each correspondence, and $P_1(t_{p_0})$ is the projection onto camera 1. If p_1 is near $P_1(t_{p_0})$, the edge between p_0 and p_1 is generated to introduce a candidate of correspondence between cameras to the graph.

between cameras $(p_0, p_1) \in S$, where p_0 and q_0 are grid points, V_0 is the set of grid points, and U_0 is the set of edges of camera 0. p_1, q_1, V_1 and U_1 are those of camera 1. S is the set of edges between cameras. A grid point p_0 has the candidates of corresponding points $t_{p_0} \in T_{p_0}$ in the projector pattern. In the case of 1C1P system, the energy of assigning corresponding point t_{p_0} to each grid point p_0 is defined as follows:

$$E(T_0) = \sum_{p_0 \in V_0} D_{p_0}(t_{p_0}) + \sum_{(p_0,q_0) \in U_0} W_{p_0 q_0}(t_p, t_q), \quad (2)$$

where $T_0 = \{t_{p_0} | p_0 \in V_0\}$. $D_{p_0}(t_{p_0})$ is the data term of assigning a candidate t_{p_0} to p_0. $W_{p_0 q_0}(t_p, t_q)$ is the regularization term of assigning candidates t_{p_0} and t_{q_0} for neighboring grid points p_0 and q_0. The data term is calculated by comparing the local pattern around the points between camera and projector images. The regularization term is zero if t_{p_0} and t_{q_0} are on the same line, otherwise it adds non-zero cost. Refer [13] for the detailed definition.

In this paper, we extend the definition of energy to the 2C1P system by introducing camera-camera constraint as follows:

$$E(T) = E(T_0) + E(T_1) + \sum_{(p_0,p_1) \in S} X_{p_0 p_1}(t_{p_0}, t_{p_1}), \quad (3)$$

where $X_{p_0 p_1}(t_{p_0}, t_{p_1})$ is the regularization term for the edges between cameras (p_0, p_1). This term is defined as follows:

$$X_{p_0 p_1}(t_{p_0}, t_{p_1}) = \begin{cases} 0 & t_{p_0} = t_{p_1} \\ \mu & \text{otherwise,} \end{cases} \quad (4)$$

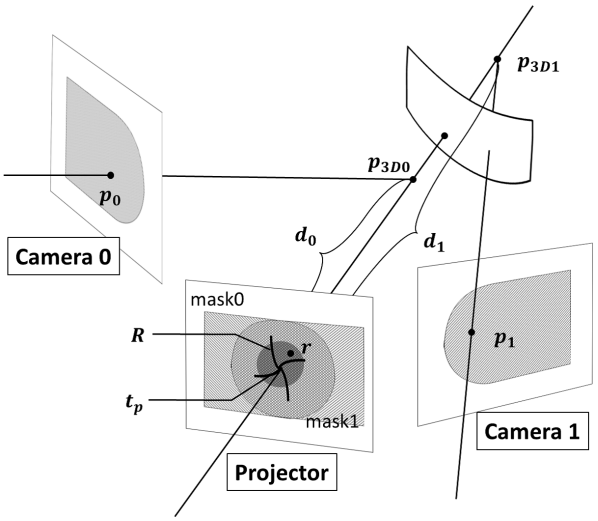


Figure 4. A grid point t_p has correspondences with points, p_0 and p_1. Two 3D points, p_{3D0} and p_{3D1}, are calculated by the two correspondences. The depth for a pixel r is calculated by averaging the depths, d_0 and d_1, if r is in the area close to t_p, which is represented by R. R is also used for creating a mask image for visibility check in Sec.4.2.

where μ is a user-defined constant. If a node p has camera-camera edges, the assignment of t_p for the node tends to be chosen from the ones with camera-camera edges, because the energy becomes large if the assignment without camera-camera edges is chosen. The energy minimization is accomplished by belief propagation [3].

4. Generating Dense Shape by Integrating Two Cameras

The grid-based stereo in the previous section gives sparse correspondences of the images of 2C1P system. The next step is to generate dense shape by using the correspondences. Since the assignments of t_p for each grid point p represents the correspondence between a camera and a projector, two sets of point clouds can be calculated from the two sets of camera-projector combination. However, the accuracy can be improved by generating a single integrated shape with two cameras and one projector. In this section, we first merge the information of two cameras into a single range image, and then optimize the range image by refining the correspondence for every pixel to generate a dense shape.

4.1. Generating Single Range Image by Merging Correspondence Information of Two Cameras

In the previous method with 1C1P system [13], a dense range image was created by interpolating the grid graph in the camera image. Now, we have two sets of grid graphs

on the 2C1P system. Since some parts of the graphs are occluded from the other camera, it is not possible to integrate the information as a range image in the coordinate system of one of the cameras. We therefore merge the information in the coordinate system of the projector by reprojecting the graphs from each camera.

Fig.4 shows the situation that a grid point t_p of the projector pattern has correspondences with points, p_0 and p_1, of both two cameras. Two 3D points, p_{3D0} and p_{3D1}, are calculated by the two correspondences, which usually do not coincide due to the error of image processing and calibration. We integrate the depths, d_0 and d_1, from the viewpoint of the projector by averaging them. To generate a dense range image, the depth d_r for a pixel r is calculated as follows:

$$d_r = \frac{1}{|R|} \sum_{(t_p,p)\in R} d(t_p, p), \quad R = R_0 \cup R_1 \tag{5}$$

$$R_k = \{(t_{p_k}, p_k) | D(r, t_{p_k}) < \tau, p_k \in V_k\}, \tag{6}$$

where $d(t_p, p)$ is the depth of the 3D points calculated by triangulation of t_p and p, and τ is a user-defined parameter to determine the neighboring grid point. Since every p_{3D} is visible from the projector, the depth information can be merged without considering occlusion.

4.2. Optimizing Range Image with 2C1P System

Next, we optimize the depth by comparing the intensities at the corresponding points. Now, the depth at a point r is d_r, and the corresponding points of the cameras are $P_k(r, d_r), (k = 0, 1)$. If we consider Δd_r, which is the small movement of d_r, the error of intensity is defined as follows:

$$E(\Delta D) = \sum_{k=0}^{1} \sum_{r} (I_{ck}(P_k(r, d_r + \Delta d_r)) - I_p(r))^2 \tag{7}$$
$$+ \gamma \sum_{r} \left((d_r + \Delta d_r) - \frac{1}{|Q(r)|} \sum_{r' \in Q(r)} (d_{r'} + \Delta d_{r'}) \right)^2,$$

where ΔD is the set of Δd_r, $I_{ck}(p)$ is the intensity of camera k at pixel p, $I_p(r)$ is the intensity of the projector pattern at pixel r, and $Q(r)$ is the set of neighboring pixels of pixel r. The first term is the data term that compares the intensities of camera and projector. The second term is the regularization term defined by the discrete Laplacian of r. d_r for each pixel is iteratively updated by adding Δd_r that minimizes the error $E(\Delta D)$.

Eq.(7) assumes that a 3D point p_{3D} calculated from the point r and its depth d_r is visible from both cameras. It is however not true in an actual situation. Therefore, the visibility check is needed in this step. We introduce a mask image corresponding to each camera in the projector coordinate system. If $P_k(r, d_r)$ for camera k is in the valid area, the data term is calculated for camera k, otherwise omitted from $E(\Delta D)$. The mask image is created during calculating the initial range image in Sec.4.1. The pixels in R_k of Eq.(6) are marked as valid for camera k.

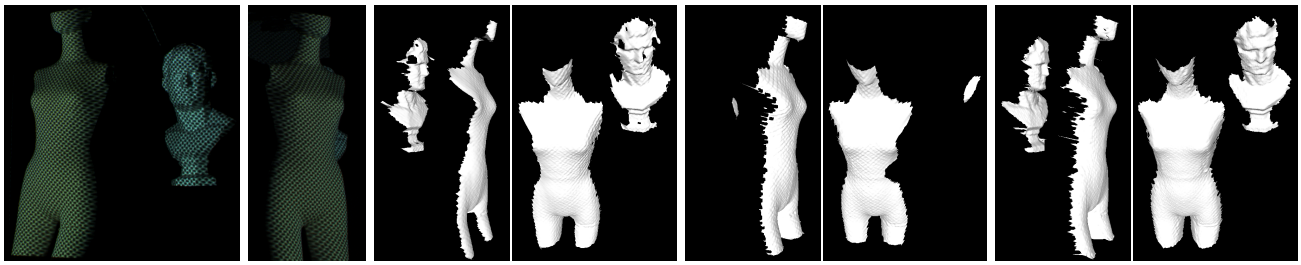
5. Experiments

We have conducted experiments to confirm the effectiveness of the proposed method. Since the contribution of this paper is the improvements of robustness and accuracy, and reducing occlusion compared to the previous method [13], we experimented the proposed method to verify these improvements. In the experiments, we used global shutter cameras of 1600 $\times$ 1200 pixels that capture images at 30 frames/second, and a liquid crystal projector 1024 $\times$ 768 pixels.

First, Fig.5 shows that the proposed method can effectively use the projected pattern and reduce the occlusion in the reconstructed shape. (a) and (b) are the input images of camera 0 and 1 with two objects. The right plaster figure is almost completely occluded in camera 1. (c), (d), and (e) are the results of reconstruction by using camera 0, camera 1, and both cameras, respectively. The left side of the mannequin is not reconstructed in (c), and the plaster figure is not in (d) due to occlusion. Since the proposed method merges the information of two cameras, the both parts are successfully reconstructed. While objects are necessary to be observed by both cameras in passive stereo systems, some of which use pattern projection, the proposed method can reconstruct the parts that are observed by only one camera because it is a structured-light system that reconstructs shapes by a pair of one projector and one camera.

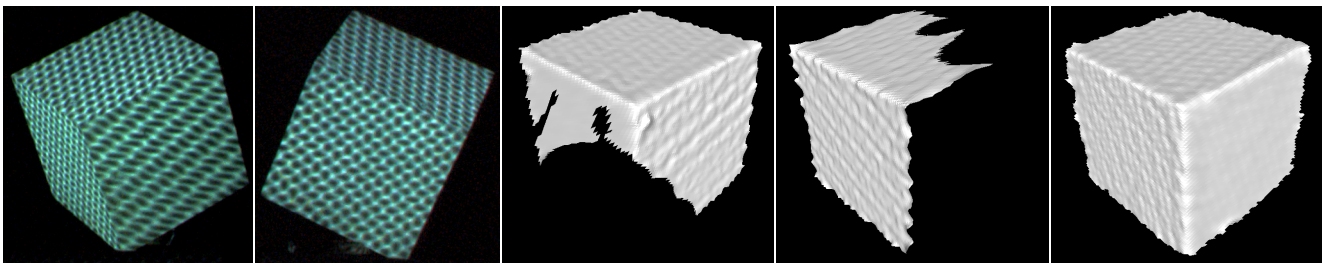
Next, we show that the robustness is improved by the proposed method. In Fig.6, a cube-shaped object is reconstructed. The cube is 0.2m on a side and the distance from the cameras is about 1.7m. (a) and (b) are the input images of two cameras. The pattern projected on the largely tilted faces is compressed and difficult to find correspondences. Therefore, the reconstructions by the 1C1P system, (c) and (d), are failed at several parts of the faces. (e) is the result of our 2C1P system. The proposed method succeeded to reconstruct all three faces of the cube. We fitted a plane to each face to evaluate the accuracy of measurements. While the RMS errors are 0.411mm and 0.398mm for the 1C1P methods with camera 0 and 1, respectively, the RMS error of our 2C1P method is 0.331mm. The accuracy is improved by using two cameras.

Next, we evaluated the accuracy by calculating the difference from the shape obtained by a close-range 3D scanner, which have better accuracy. Fig.7 shows the difference by color. The object is about at 2.0m from the cameras. (a) and (b) are the results from the 1C1P and 2C1P meth-



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

Figure 5. (a) and (b) are the input images of camera 0 and 1 with two objects. The right plaster figure is almost completely occluded in camera 1. (c), (d), and (e) are the results of reconstruction by using camera 0, camera 1, and both cameras, respectively.



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

Figure 6. (a) and (b) are the input images of two cameras. (c) and (d) are the reconstructions by the 1C1P system. (e) is the result by our 2C1P system.

ods, respectively. It can be recognized that the difference in (b) is smaller than that in (a) by color. Since the RMS differences are 2.63mm and 2.22mm for 1C1P and 2C1P methods, respectively, it indicates the accuracy is improved by the proposed method.

Next, Fig.8 shows an example of the extreme case that the projector and cameras are at degenerated position. We tested with two camera arrangements shown in Fig.8 (1). The camera arrangement (a) is configured so that all the epipolar lines and grid lines to not be parallel to avoid degenerate position, whereas, the camera arrangement (b) is at a degenerated position that one of the camera was set to be parallel. Since the epipolar plane between camera 0 and the projector is parallel to one of the wave line direction, the wave line does not curve regardless of the shape of the object. It means the connectivity of grid points given by the wave line has no information, which causes the failure of 3D reconstruction by the 1C1P system [13].

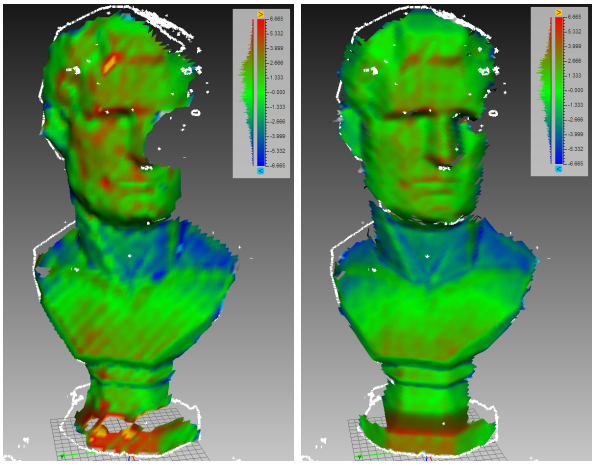
The input images are shown in Fig.8(2) and (3). Since the horizontal wave line is parallel to the epipolar line between camera 0 and the projector in the configuration (b), the pattern observed by camera 0 does not curve. In this case, the connectivity obtained by the horizontal pattern does not have useful information, because it is trivial by the epipolar geometry. Fig.8(4), (5) and (6) show the reconstructed results from camera 0, camera 1, and from both cameras, respectively. All the cases succeeded to reconstruct the shape of a sphere in the setup (a). While camera 0 in the setup (b) failed, the proposed method successfully reconstructed the shape.

Table 1. The number of correct/incorrect correspondences by using one camera or two cameras.

	Num of grid points	1C1P method		2C1P method	
		correct	incorrect	correct	incorrect
(a)	432	432	0	432	0
(b)	540	448	92	536	4

Table 1 shows the statical result on detecting correct correspondences of grid points for the cases in Fig.8. The numbers of grid points are the sum of points detected on camera 0 and 1. In the setup (a), the reconstructions by both 1C1P and 2C1P systems find correct correspondences from all grid points. In the setup (b), many incorrect correspondences by the 1C1P method are chosen because of the failure of camera 0. The number of incorrect correspondences is significantly reduced by the 2C1P method, which indicates that the correct correspondences are found even in camera 0 of degenerated position.

Finally, Fig.9 shows an example of capturing a person in motion. (a) and (b) are three frames of the input video of camera 0 and 1, respectively. (c), (d), and (e) are the results of reconstruction by using camera 0, camera 1, and both cameras. In some cases, the 1C1P method fails to find the



(a) (b)

Figure 7. The differences from the shape obtained by a close-range 3D scanner are calculated. (a) and (b) are the results by the 1C1P and 2C1P methods, respectively.

correct correspondence because of the inappropriate configuration of the pattern. Wrong correspondences occur at similar positions in different frames as shown by red circles in (c) and (d). Since the wrong correspondences are reduced by the 2C1P method, it shows the proposed method improves the robustness to find the correct correspondence.

6. Conclusion

In this paper, we proposed an active 3D reconstruction with two cameras and one projector (2C1P) system. The system reconstructs the shapes from a single frame of each camera by finding the correspondence between the cameras and the projector based on projected wave grid pattern. We introduced camera-camera constraint, which is represented as an edge between the grid graphs detected in camera images, to extend the one-shot 1C1P method with the wave grid pattern. Since the corresponding information of two cameras is merged and optimized, it was shown in the experiments that the proposed method contributes to reducing occlusion, and improving robustness and accuracy. In future work, we plan to extend the proposed method to capture the entire shape of an 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] E. de Aguiar, C. Stoll, C. Theobalt, N. Ahmed, H.-P. Seidel, and S. Thrun. Performance capture from sparse multi-view video. In *ACM Trans. Graphics (Proc. SIGGRAPH)*, volume 27, page 98, 2008. 1

[3] P. Felzenszwalb and D. Huttenlocher. Efficient belief propagation for early vision. *IJCV*, 70:41–54, 2006. 4

[4] R. Furukawa, R. Sagawa, H. Kawasaki, K. Sakashita, Y. Yagi, and N. Asada. Entire shape acquisition technique using multiple projectors and cameras with parallel pattern projection. *IPSJ Transactions on Computer Vision and Applications*, 4:40–52, Mar. 2012. 2

[5] 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, 3

[6] 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

[7] 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

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

[9] 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

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

[11] R. Sagawa, H. Kawasaki, R. Furukawa, and S. Kiyota. Dense one-shot 3D reconstruction by detecting continuous regions with parallel line projection. In *ICCV*, pages 1911–1918, 2011. 1

[12] 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. 2, 3

[13] R. Sagawa, K. Sakashita, N. Kasuya, H. Kawasaki, R. Furukawa, and Y. Yagi. Grid-based active stereo with single-colored wave pattern for dense one-shot 3D scan. In *Proc. 2012 Second Joint 3DIM/3DPVT Conference*, pages 363–370, Zurich, Switzerland, Oct. 2012. 2, 4, 5, 6

[14] J. Salvi, J. Batlle, and E. M. Mouaddib. A robust-coded pattern projection for dynamic 3D scene measurement. *Pattern Recognition*, 19(11):1055–1065, 1998. 2

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

[16] 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

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

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

[19] A. O. Ulusoy, F. Calakli, and G. Taubin. One-shot scanning using de bruijn spaced grids. In *The 7th IEEE Conf. 3DIM*, pages 1786–1792, 2009. 1, 2

[20] 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

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

[22] 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

[23] 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

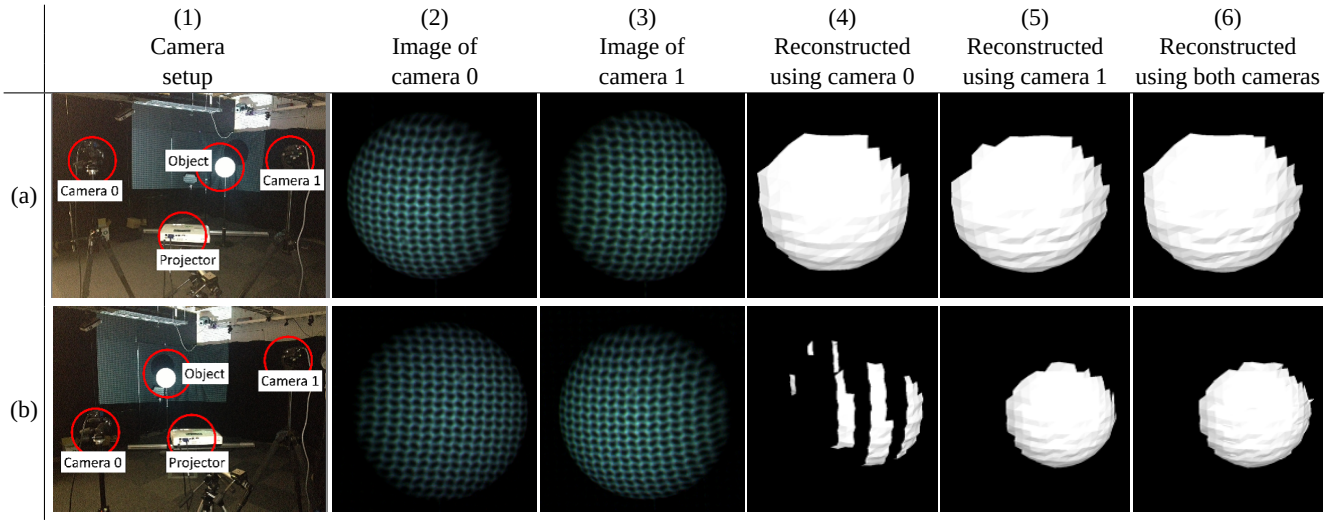
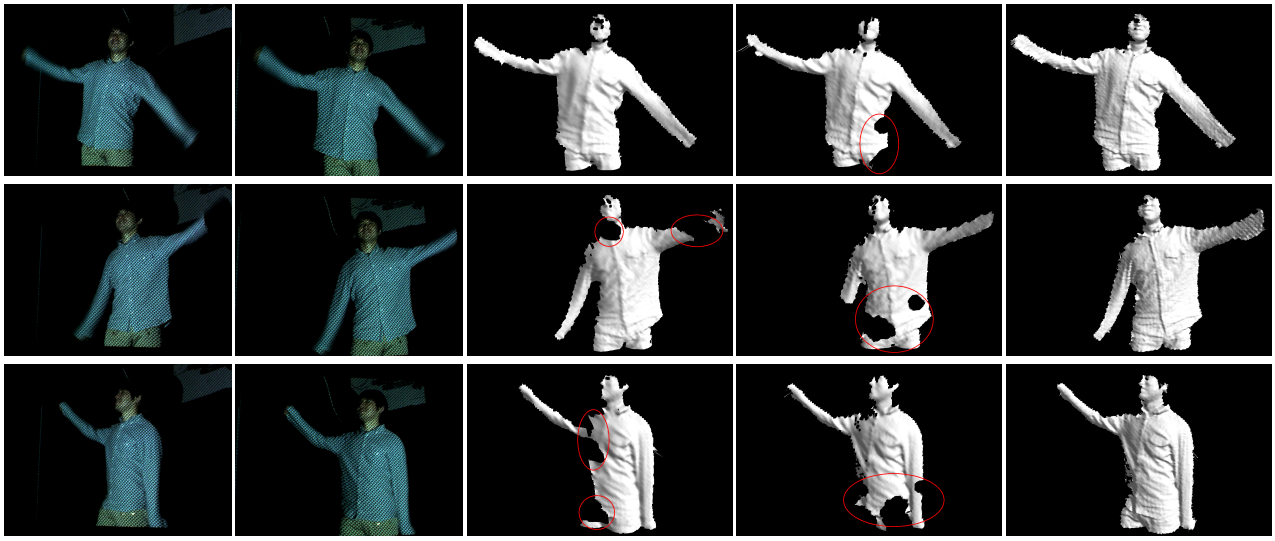


Figure 8. The proposed method (2C1P) is compared to 1C1P system with two different setups, (a) and (b). (a) is configured so that all the epipolar lines and wave lines to not be parallel to avoid degenerate position, whereas, (b) is a degenerated setup that one of the camera was set to be parallel. Therefore, the pattern observed by camera 0 does not curve in (b)-(2). All the cases succeeded to reconstruct the shape of a sphere in the setup (a). While camera 0 in the setup (b) failed, the proposed method finds correct correspondences.



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

Figure 9. An example of capturing a person in motion: (a) and (b) are three frames of the input video of camera 0 and 1, respectively. (c), (d), and (e) are the results of reconstruction by using camera 0, camera 1, and both cameras. Red circles indicate wrong correspondences.