One-shot Entire Shape Scanning by Utilizing Multiple Projector-Camera Constraints of Grid Patterns

Nozomu Kasuya
Kagoshima University
Kagoshima, 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

This paper proposes a method to reconstruct the entire shape of moving objects by using multiple cameras and projectors. The projectors simultaneously cast static grid patterns of wave lines. Each of the projected patterns is a single-colored pattern of either red, green, or blue. Those patterns can be decomposed stably, compared to multi-colored patterns. For the 3D reconstruction algorithm, one-shot reconstruction with wave grid pattern is extended for entire-shape acquisition, so that the correspondences between the adjacent devices can be used as additional constraints to reduce shape errors. Finally, multiple shapes obtained from the different views are merged into a single polygon mesh model using estimated normal information for each vertex.

1. Introduction

In recent years, as practical and reliable active 3D scanning devices develop rapidly, strong demands on capturing motions of humans or animals as a series of entire 3D shapes is emerging [7]. Such 3D data can be applied to gesture recognition, markerless motion capture, digital fashion, and analysis of interaction between multiple humans or animals.

An important technical issue for the entire-shape scan of a moving object is that a general 3D sensor can only capture surfaces visible from one direction during one measurement; thus, acquiring an entire shape of an object is difficult. One approach that has been widely used is aligning a large number of cameras around the object, and reconstruct the 3D shape with shape-from-silhouette [3, 13] . These passive measurement systems have achieved major successes; however, setting up those systems is complicated since they need a large number of cameras with broad networks and PCs, and the calculation time is huge.

Active entire-shape measurement systems could be alternatives to overcome the above problems of the passive methods. However, since active methods use light sources, interferences between different illuminations become a major problem. Several methods that have been proposed until now can be largely separated into two types; one is to use different colors (i.e., wave-lengths), the other is high-frequency switching between multiple light sources with synchronized camera capturing. Since the timing of capturing for multiple light sources are different, temporal interpolation is needed for the integration of the shapes [20, 2], which may limit the extent of the applicability, the latter method is not suitable to scan fast-moving objects. Therefore, we take the former option in our method; however, the former method also has problems.

If each of the light sources project a light pattern with multiple colors, it is difficult to decompose the light patterns from the captured image in which multiple patterns are overlapped on a same surface; thus, devices that just use single-colored pattern is suitable for the system, such as Kinect. However, just aligning multiple devices with such a property also has problems, such as inconsistency between shapes measured by difference devices that come from calibration errors of intrinsic or extrinsic parameters.

An approach to deal with this problem is using correspondences between multiple patterns projected from multiple light projectors. As far as we know, there has been no method that performs this method with single-colored patterns. The method proposed by Furukawa et al. is related to

this approach [5]; however, it sometimes fails to decompose the projected patterns because each pattern has two colors.

In this paper, we propose an active entire-shape measurement system in which single-colored patterns are projected, and correspondences between multiple patterns are used. To achieve this, we extend one of the spatial-encoding methods proposed by Sagawa et al. [17], which reconstructs shapes by projecting a static grid pattern of wave lines, to entire-shape acquisition. In the system, each projector emits a single-colored waved-grid pattern with a color of either red, green, or blue, and the multiple grid-patterns are overlapped. The target is captured from multiple cameras. Similarly to Sagawa et al. [17], the reconstruction is achieved by belief propagation (BP) on grid structure, except that consistencies between multiple devices to suppress errors of the entire-shape are considered. The contributions of the proposed method is as follows:

- Decomposition of the pattern illuminated from different projectors becomes stable since single-colored patterns with RGB components are used.
- In the process of deciding stereo correspondences with grid-based BP, correspondences between multiple patterns from different projectors captured from a single camera is used to stabilize the BP process.
- In the process of entire-shape reconstruction, correspondences between points on different depth images generated with viewpoints of different projectors are synthesized, and the errors for the corresponding points are minimized to improve the shape accuracy.
- Since the normal information for all the vertex is obtained simultaneously, a single and closed polygon manifold mesh can be efficiently generated with Poisson reconstruction methods(e.g., [10]).

From the above contributions, the complexity of the system is drastically reduced, the errors between multiple devices can be globally minimized to obtain consistent entire-shape, and dense and accurate reconstruction is achieved because errors of pattern decomposition is reduced.

2. Related Work

To capture the entire shape of moving objects, a silhouette based shape reconstruction method using a number of camera has been widely researched and developed [3]. Although the system have been constructed for various purposes over the world, there are several drawbacks that still exist for the method, such as complicated and laborious task for the set-up and the calibration, low quality of silhouette based reconstruction and huge computational cost.

One simple solution is to use an active light source to decrease the number of necessary devices. However, considering the use of multiple lights for capturing the entire shape, several light sources are projected onto the same object and such severe interference of the multiple patterns makes it impossible to reconstruct the shape with original active stereo algorithm. To solve the issue, two main approaches are proposed. The first approach is quickly switching the light sources and the other is using different wavelengths or modulations for each light sources. In terms of light switching approach, it is reported that 60 fps is realized for the fast moving human as a target [20, 2]. However, since the technique essentially requires integration of multiple shapes to be captured at different times, applicable conditions are limited. Further, synchronous switching of a number of light sources with multiple cameras is difficult.

For the technique using different wavelengths or modulations for each light source, if many wavelengths are overlapped together at the same place, it is difficult to decompose them into the original component. Therefore, reducing the number of wavelength for each light source is the key for the approach. For example, since TOF method usually uses a laser with single wavelength, it is possible to use multiple of them to capture a wide range [11]. One problem is that the resolution of one-shot TOF scanner is low [14].

Another example is an active stereo method, which only uses a limited number of wavelength for scan [15]. In the method, six sets of Kinect are used for capture the wide range of the scene. Since they use the same wavelength for all the devices, it sometimes fails to capture the object. Further, for both TOF and active stereo method, since shapes are reconstructed independently, there remains severe inconsistency during the integration of multiple shapes if calibration has some errors.

To solve such inconsistency, instead of independent reconstruction of each shape, simultaneous method for an entire shape scan is proposed [3]. In the method, multiple parallel lines are projected from each projector and each camera captures multiple patterns with their intersections. Since the intersection has some constraint on their original lines, the shape is reconstructed by using the constraints. One problem of the method is that it uses two colors for each projector, and thus, it is overlapped on the same area of the object surfaces, resulting in erroneous reconstruction.

Recently, active one-shot method is widely researched and is being developed because of a strong demand to scan moving objects [18, 21, 12, 1, 9, 16, 19]. Especially, single-color based one-shot scan is intensively researched since it is robust on image processing and can simplify the construction of light source [1, 17]. In this paper, we extend the single color one-shot scanning technique for the purpose of entire shape acquisition by using multiple sets of them.

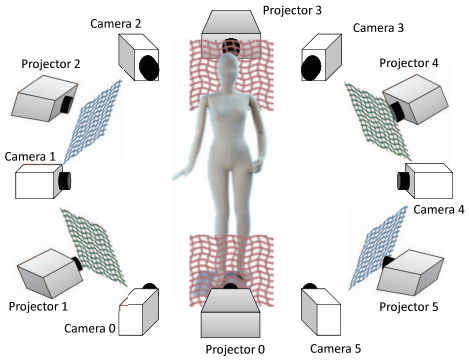


Figure 1. The proposed system captures entire shape by using multiple cameras and projectors.

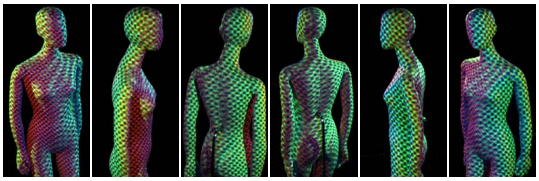


Figure 2. An example of input images

3. Overview

The proposed system consists of multiple cameras and projectors as shown in Fig.1. Each projector casts a static wave-grid pattern proposed in [17]. The cameras capture images synchronously, but no synchronization between a camera and projectors is required because the projected pattern does not change. A 3D shape at the moment of image acquisition is reconstructed from the single frame of the cameras.

The proposed method reconstructs a shape basically using a pair of a camera and projector neighboring each other. Since a camera shares the field of view with multiple projectors in the system, multiple patterns from different projectors can be captured by a camera. In doing so, it is necessary to recognize the correspondence between projector and detected pattern. While lines of multiple colors are projected for reconstruction from each projector in [6], the reconstruction with wave-grid pattern uses single-colored pattern for a projector. In this paper, we can therefore assign different colors for the projectors. Because an off-the-shelf video projector can cast RGB colors, we locate the projectors so that neighboring projectors do not use the same colors as shown in Fig.1. Fig.2 shows an example of input images of the system with 6 cameras and 6 projectors. The details of grid detection from the images are described in Sec.4.

The second step is grid-based active stereo with multiple cameras and projectors, which finds the correspondence of intersection points of the grid patterns (grid points) between cameras and projectors. Our method is based on the method proposed in [17], which uses one camera and one projector. If we use multiple cameras and projectors, additional constraints on the correspondence between two cameras, or two projectors can be introduced. The details of grid-based stereo are explained in Sec.5.

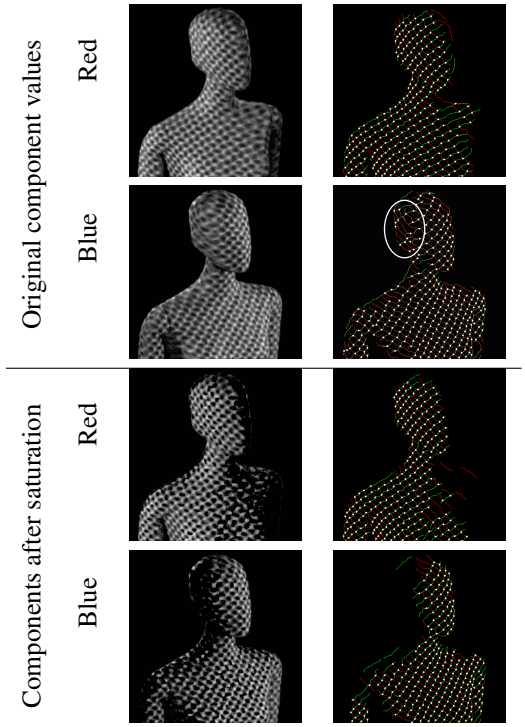


Figure 3. The extraction of red and blue patterns is improved by saturating the colors.

The third step is optimization of correspondence by using all pixels of cameras and projectors. The correspondence given by grid-based stereo is used as initial value and interpolated by using all pixels to obtain dense 3D shape. The problem becomes multi-view stereo with cameras and projectors, but it is simplified due to pattern projection because initial guess and visibility check are given by grid-based stereo. The detail is described in Sec.6.

4. Detecting Grid Patterns from Multiple Projectors

A grid pattern consists of wave lines of two directions that are perpendicular to each other. The projected patterns are first detected as curves in a captured image, and the grid points are calculated from the intersection points of the curves. Since the connection of grid points is used for finding correspondences, it is necessary to discriminate wave lines of different directions. We use the method of curve detection based on belief propagation proposed by Sagawa et al. [16]. The method separately detects the lines of two directions even if they have the same color.

Since multiple patterns are observed in a camera image, it is required to recognize which projector casts the pattern.

We use the colors for this purpose. The projected pattern can be red, green, or blue because commercial video projectors are used in our experimental system. The RGB spectra of both camera and projector usually overlap each other and it causes the crosstalk between the color components of the captured image. In Fig.3, we want to extract the patterns projected as red and blue patterns from projectors in this image, while the projectors of green pattern are not used because they are far from the camera to find its correspondence. If we detect lines by using red and blue components of the input image, the results are affected by the green pattern. The green lines are detected in the result of blue component at the side of the head (white circled).

To determine which projector a pattern is emitted from, we saturate the colors before grid detection as follows:

$$\begin{aligned}(h, s, v) &= \text{RGB2HSV}(r, g, b)\\ (r', g', b') &= \text{HSV2RGB}(h, 1, v),\end{aligned} \tag{1}$$

where RGB2HSV and HSV2RGB are functions that converts between RGB and HSV color space, and colors are represented in the range of $[0, 1]$. By saturating the colors, minor components are suppressed. It is avoided that green lines are detected in blue components.

5. Grid-based Active Multi-view Stereo

This section describes the proposed method of grid-based active stereo. It is based on a method proposed by Sagawa et al. [17]. The basic system has a single projector and a single camera (1C1P). If two cameras can be used, a method that extends the basic method is proposed in [8] to improve the robustness and accuracy by introducing additional constraint between cameras. Since we have multiple cameras and projectors in this paper, more constraints can be added to the scheme for finding correspondence on the graph created by the grid patterns.

5.1. Finding Correspondence with Wave Grid Pattern between Camera and Projector

If the system is calibrated, the epipolar line in the projector image that corresponds to a grid point in the camera image can be calculated. The true corresponding point of the grid point on the projector image is the one from intersection points on the epipolar line. First, the method collects all intersection points on the epipolar line as the candidates of correspondence. Since the grid points are connected by the detected curves, the grid points are regarded as nodes of a graph. The problem of finding correspondence becomes the one that chooses the best candidate assigned to each node.

Now, the grid graph detected in camera i consists of nodes $p_i \in V_i$, edges by curve detection $(p_i, q_i) \in U_i$, where p_i and q_i are grid points, V_i is the set of grid points,

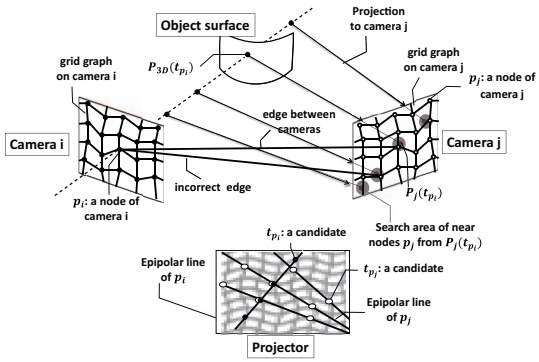


Figure 4. The edges between p_i and p_j is generated if they satisfy geometrical constraint.

and U_i is the set of edges. A grid point p_i has the candidates of corresponding points $t_{p_i} \in T_{p_i}$ in the projector pattern. In the case of 1C1P system, the energy of assigning corresponding point t_{p_i} to each grid point p_i is defined in as follows:

$$E(T_i) = \sum_{p_i \in V_i} D_{p_i}(t_{p_i}) + \sum_{(p_i, q_i) \in U_i} W_{p_i q_i}(t_p, t_q), \tag{2}$$

where $T_i = \{t_{p_i} | p_i \in V_i\}$. $D_{p_i}(t_{p_i})$ is the data term of assigning a candidate t_{p_i} to p_i. $W_{p_i q_i}(t_p, t_q)$ is the regularization term of assigning candidates t_{p_i} and t_{q_i} for neighboring grid points p_i and q_i. 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_i} and t_{q_i} are on the same line; otherwise it adds non-zero cost. Refer [17] for the detailed definition. The assignments of correspondence are determined by minimizing the energy accomplished by belief propagation (BP) [4].

5.2. Finding Correspondence with Constraints between Cameras

If multiple cameras can observe the projected pattern emitted by a projector, the assignments of correspondence between projector and camera must satisfy the geometrical constraint between the two cameras, which is introduced as the edges that connect graphs of two cameras. This approach is proposed in [8] and briefly explained here.

Fig.4 shows how to generate edges between two graphs. First, we detect the wave line in the camera images and create the grid graphs. Next, let us determine the corresponding point in the projector pattern of a node p_i of camera i, which is a grid point where two lines intersect. The candidates of the corresponding points $t_{p_i} \in T_{p_i}$ are the intersection points of the pattern on the epipolar line of p_i in the projector image, where T_{p_i} are the set of the candidates for the node p_i. If we assume the correspondence of p_i and t_{p_i}, the 3D coordinates $P_{3D}(t_{p_i})$ for the nodes p_i are calculated by triangulation between camera i and the projector. Next, the projection of the 3D points $P_{3D}(t_{p_i})$ onto the image

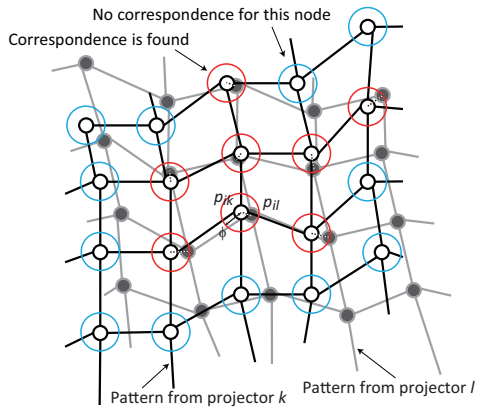


Figure 5. The edges between two graphs in a camera image are generated.

of camera j is $P_j(t_{p_i})$ as shown in Fig.4. If the node p_j of camera j is close to $P_j(t_{p_i})$, p_i and p_j can be corresponding points. Four $P_{3D}(t_{p_i})$ are projected onto camera j in Fig.4. Since the leftmost $P_j(t_{p_i})$ has no nodes in the search area, no candidate of correspondence is found. While the rightmost one has a node p_j in the search area, the node does not have the same candidate t_{p_i} in T_{p_j}. Since the middle two projections satisfy the above condition, their nodes are connected to p_i. Once the edges between two cameras connect their graphs, they become a single graph, which enables us to simultaneously optimize the correspondences search of two cameras.

By using the constraint between cameras, the energy of the edges between cameras is defined as follows:

$$X_{p_i p_j}(t_{p_i}, t_{p_j}) = \begin{cases} 0 & t_{p_i} = t_{p_j} \\ \mu & \text{otherwise,} \end{cases} \tag{3}$$

where μ is a user-defined constant.

5.3. Finding Correspondence with Constraints between Projectors

The proposed system uses multiple projectors. If the projected area overlapped with each other, the correspondence between projectors can be detected by using camera images. Fig.5 shows a situation that two patterns are overlapped in a camera images. If two grid points of different patterns are projected on the same pixel of the camera, it means that the two points of projectors corresponds each other. In that case, the two points have the same depth from the camera. Since it is not so frequent case that two points are projected onto the exact same pixel, we determine the corresponding point $p_{ik} \in V_{ik}$ of camera i for the projector k the by finding $p_{il} \in V_{il}$ of of camera i for the projector l that satisfy the following condition:

$$D(p_{ik}, p_{il}) < \phi, \tag{4}$$

where $D(a, b)$ is the distance between points a and b, and ϕ is the radius of search area around p_{ik}.

The corresponding nodes of two graphs are connected as the dotted lines shown in Fig.5. Therefore, the two graphs become one and the assignments are simultaneously optimized by minimizing the energy. The energy of the edges of projector-projector correspondence is defined as follows:

$$Z_{p_{ik}p_{il}}(t_{p_{ik}}, t_{p_{il}}) = \tau |d_i(P_{3D}(t_{p_{ik}})) - d_i(P_{3D}(t_{p_{il}}))|, \tag{5}$$

where $d_i(P_{3D})$ is the depth of the 3D point P_{3D} in the coordinate of camera i, and τ is a user-defined weight.

The total energy with multiple cameras and projectors is defined by the following equation:

$$E(T) = \sum_i \sum_{k \in A_p(i)} E(T_{ik}) \tag{6}$$

$$+ \sum_k \sum_{i \in A_c(k), j \in A_c(k)} \left(\sum_{(p_{ik}, p_{jk}) \in S_{ijk}} X_{p_{ik}p_{jk}}(t_{p_{ik}}, t_{p_{jk}}) \right)$$

$$+ \sum_i \sum_{k \in A_p(i), l \in A_p(i)} \left(\sum_{(p_{ik}, p_{il}) \in Q_{ikl}} Z_{p_{ik}p_{il}}(t_{p_{ik}}, t_{p_{il}}) \right),$$

where $A_p(i)$ is the set of projectors that share the field of view with camera i, $A_c(k)$ is the set of cameras that share the field of view with projector k. S_{ijk} is the set of edges between cameras i and j given by the pattern of projector k. Q_{ikl} is the set of edges between projectors k and l in the image of camera i.

6. Generating Dense Shape by Integrating All Cameras and Projectors

The grid-based stereo in the previous section gives sparse correspondences for the grid points of projected patterns. The next step is to generate dense correspondences by using all pixels and a single dense intricate shape by using all cameras and projectors. To integrate the information of all cameras, it is necessary to check the visibility of each 3D points from cameras, which is a common problem of multi-view stereo. In our proposed system, however, the issue can be simplified because the visibility from a projector is given from the projected pattern.

The proposed integration consists of three steps. The first step is to generate dense range images from the view point of projectors. The second step is to optimize the range images to minimize the errors with camera images and the discrepancy between range images. Third, all range images are merged into a single mesh model.

6.1. Generating Range Image for Each Projector

Once a grid pattern is detected as that of the projectors, it is obvious that the detected points are visible from the projector. Because the correspondence between camera and

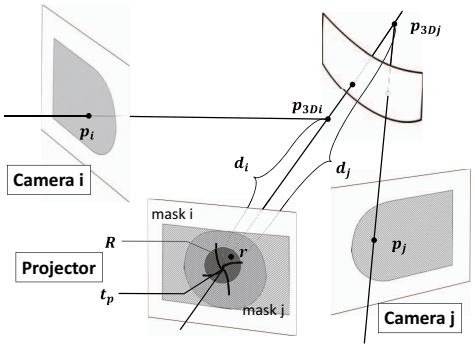


Figure 6. The range image from the viewpoint of projector is generated.

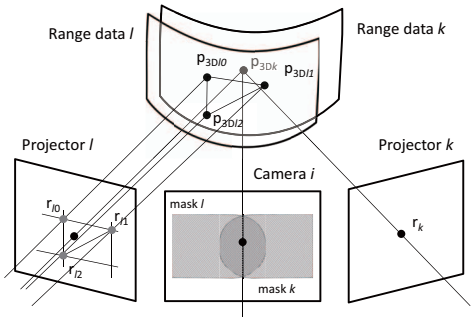


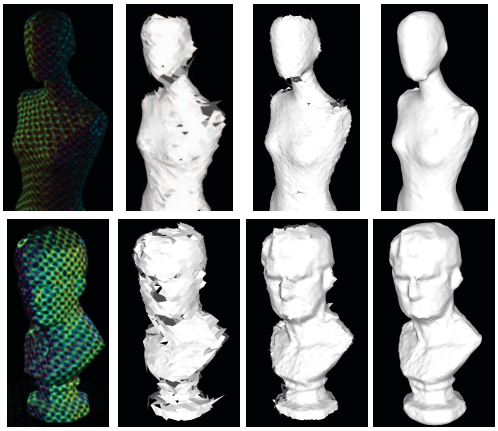
Figure 7. The difference between two range data is minimized by finding the correspondence between projectors.

projector is roughly known by grid-based stereo, a range image from the view point of the projector can be created. If the pattern from a projector is observed by multiple cameras, it is easy to merge the information as a range image without considering the visibility. This approach is proposed in [8] and briefly explained here.

Fig.6 shows the situation that a grid point t_p of the projector pattern has correspondences with points, p_i and p_j, of both two cameras. Two 3D points, p_{3Di} and p_{3Dj}, are calculated by the two correspondences, which usually do not coincide due to the errors of image processing and calibration. We integrate the depths, d_i and d_j, 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 by weighted average of depths of grid points inside the circle of radius R around r. The mask image is created during the calculation of the range image. The pixels in R is marked as valid for camera i.

6.2. Simultaneous Optimization of Multiple Range Images

Next, we optimize the depths of all range images by minimizing the energy. The energy proposed in [8] consists of the data term and regularization term. The data term is calculated by the difference of intensities between the camera and the projector, and the regularization term is defined by using the curvature around each vertex of the mesh model. Now, we have multiple range images and optimize them simultaneously. If two range images are overlapped with each other, the shapes should coincide, which can be used additional constraint to optimize the depths.



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

Figure 8. 3D reconstruction of two static objects:(a) input images, (b) grid-based reconstruction, (c) range data of multiple projectors, and (d) integrated shape.

Fig.7 shows a situation that two range data of projector k and l overlap. A 3D point p_{3Dk} is calculated from a point r_k of projector k. The point overlaps with projector l if the projection of p_{3Dk} is in the mask for projector l. Then, p_{3Dk} is projected onto the image of projector l. If it is inside of a triangle formed by three points, r_{l0}, r_{l1} and r_{l2}, they become the corresponding points.

Now, the depth at a point r is d_r, and we consider Δd_r, which is the small movement of d_r, which is used to update the depth during iterative minimization. The energy is defined by using Δd_r as follows:

$$E(\Delta D) = \sum_k E_I + \alpha \sum_k E_S + \beta \sum_i \sum_{k,l \in A_p(i)} E_P, \quad (7)$$

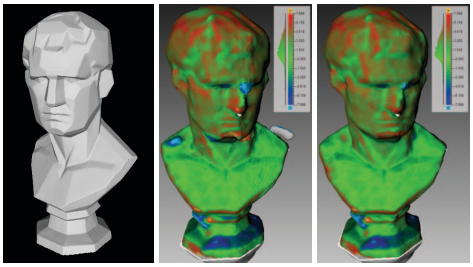
$$E_P = \sum_{r_k} \sum_{r_{l_n} \in G(r_k)} (P_{3Dk}(\Delta d_{r_k}) - P_{3Dl_n}(\Delta d_{r_{l_n}}))^2,$$

where ΔD is the set of Δd_r. E_I and E_S are the data and regularization terms, respectively. Refer [8] for the detailed explanation. E_P represents the constraint between range images. $G(r_k)$ is the function to find the corresponding point r_{l_n} of r_k. $P_{3D}(\Delta d_r)$ means the 3D point after p_{3D} moves Δd_r along the line of sight. d_r for each pixel is iteratively updated by adding Δd_r that minimizes the error $E(\Delta D)$ in non-linear minimization manner.

Finally, a single mesh model is created from the set of multiple range data. In this paper, we apply Poisson reconstruction proposed in [10] to merge range data.

7. Experiments

We have conducted experiments to confirm the effectiveness of the proposed method. The contribution of this



(a) (b) (c)

Figure 9. The differences of the basic method (b) and the proposed method (c) from the reference shape (a) are shown by color.

Table 1. The numbers of incorrect correspondences are compared by changing the constraint used in the grid-based stereo: (a) mannequin, (b) plaster figure.

	Total	Incorrect matches			
	Num.	Basic	w/ X	w/ Z	w/ X,Z
(a)	4950	271	100	101	69
(b)	2803	308	91	150	75

Table 2. The difference from the reference shape is evaluated by changing the constraint used. The unit is millimeter.

Basic	w/ X	w/ X,Z	w/ X,Z,E_p
2.960	2.867	2.694	2.686

paper is to improve robustness and accuracy by introducing additional constraints to the framework of grid-based stereo, and we show that the computational time is sufficiently low to process many frames as one of advantages of using the active approach. In the experiments, we used six cameras of 1600 $\times$ 1200 pixels that capture images at 30 frames/second, and six liquid crystal projector 1024 $\times$ 768 pixels, which are placed around the target objects. The distance of objects from cameras and projectors is about 2.0m in the experimental system.

First, we captured static objects and evaluated the robustness and accuracy of finding correspondence. Fig.8 shows the results of two static objects. (a) is one of input images. (b) is the result of grid-based stereo, and (c) is the dense range data after optimization. (d) is the merging result of the range data.

We evaluate the robustness of finding corresponding grid points by adding new constraints proposed in this paper. The number of incorrect matches is counted for the results obtained by changing the constraint used in the grid-based stereo. Table 1 shows the results of two objects shown in Fig.8. The basic method does not use the both constraints, $X_{p_{ik}p_{jk}}$ and $Z_{p_{ik}p_{il}}$ in Eq.(6). We tested to use either X or Z to compare with the proposed method that uses the both X and Z. The number of incorrect matches is clearly reduced by the additional constraints. The number of incorrect matches on the proposed method is about one fourth of the basic method.

Next, we evaluated the accuracy by calculating the difference from the reference shape obtained by a close-range 3D scanner, which is shown in Fig.9(a). The height of the figure is about 0.6m. The differences are shown by color in the basic method (b) and the proposed method (c). The green area, which indicates that the difference is small, is increased in (c) compared to (b). The average difference is summarized in Table 2. E_p means the constraint used in Eq.(7) of the optimization. The accuracy is improved by adding the constraints.

Table 3. Computational time for reconstruction in seconds.

	Mannequin	Plaster figure
Grid detection	2.518	2.463
Grid-based stereo	0.517	0.381
Optimization	0.256	0.228
Integration	4.205	4.159
Total	7.498	7.233

One of the advantage of the proposed method is the two-step approach for finding correspondence that rough estimation is done by grid-based stereo and dense shape is obtained by optimization. Moreover, the algorithm is easy to be parallelized by using GPU. The computational time for each step of the algorithm is summarized in Table 3. We used a PC with Intel Xeon 3.07GHz and NVIDIA Quadro4000. The numbers of vertices of the reconstructed models are about 22K and 54K, respectively. Except for integration, the computation is mostly accomplished by GPU. Since the time for a frame is less than 10 seconds, the models for a long image sequence can be generated in a reasonable time. The major parts of the time are grid detection and integration. Since the grid detection is calculated for 12 pairs of camera and projector in the experiment, it is possible to speed up by further parallelization. We used the implementation provided by Kazhdan [10] for integration. One of future work is to develop a method of integration specialized to the proposed method to speed up.

Finally, Fig.10 shows an example of capturing a person in motion. They are eight frames out of 60 frames captured at 30 frames/second. The entire shape of the person is successfully reconstructed, and it is sufficiently dense and accurate to represent the details of the cloth.

8. Conclusion

In this paper, we proposed a one-shot active 3D reconstruction method for capturing the entire shapes of moving objects. The system reconstructs the single shape, which is projected by multiple projectors and is captured by multiple cameras. This is done by finding the correspondences between the cameras and the projectors and applying our original simultaneous reconstruction algorithm. For stable and robust image processing, we use just single color for each pattern, which acquires reliable decomposition from multiple overlapped patterns on the same object. Another contribution of the paper is an efficient representation of the correspondences between multiple cameras and projectors

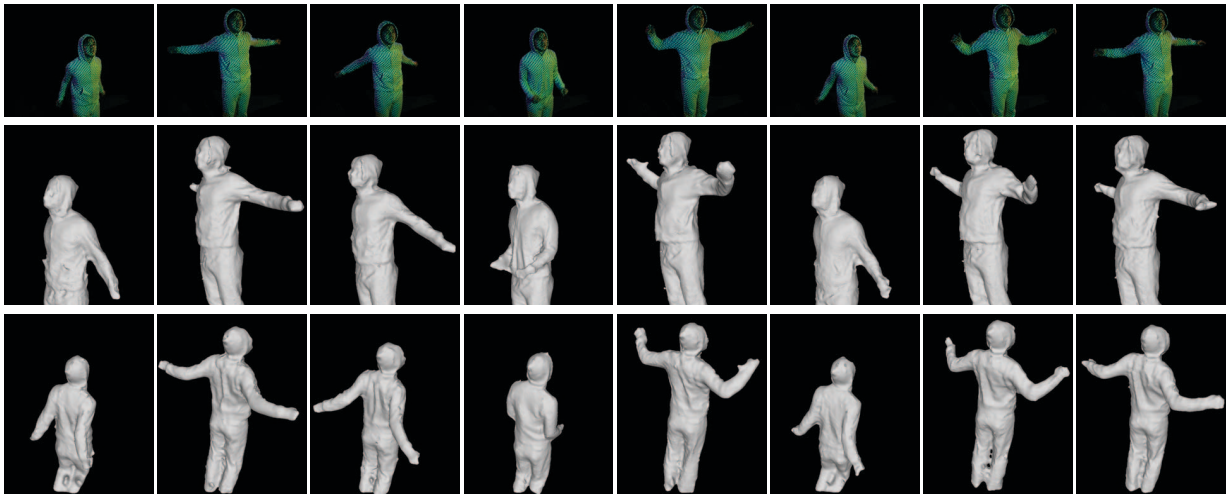


Figure 10. 3D reconstruction of a jumping person.

using graph structure, which is suitable to apply state-of-the art optimization techniques, such as BP or GraphCut. In the experiments, consistent shapes are reconstructed with our technique, whereas independent reconstruction results showed several gaps between shapes. In the future, we plan to extend the proposed method to achieve real-time processing using GPU and reconstruct a single surface directly.

Acknowledgment

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

References

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

[2] Q. D. Chenglei Wu, Yebin Liu and B. Wilburn. Fusing multi-view and photometric stereo for 3D reconstruction under uncalibrated illumination. IEEE Transactions on Visualization and Computer Graphics, 17(8):1082–1095, 2011. 1, 2

[3] G. K. M. Cheung, T. Kanade, J.-Y. Bouguet, and M. Holler. A real time system for robust 3D voxel reconstruction of human motions. In CVPR'00, pages 2714–2720, 2000. 1, 2

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

[5] R. Furukawa, R. Sagawa, H. Kawasaki, K. Sakashita, Y. Yagi, and N. Asada. One-shot entire shape acquisition method using multiple projectors and cameras. In 4th Pacific-Rim Symposium on Image and Video Technology, pages 107–114. IEEE Computer Society, 2010. 2

[6] 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. 3

[7] J. Gall, C. Stoll, E. de Aguiar, C. Theobalt, B. Rosenhahn, and H.-P. Seidel. Motion capture using joint skeleton tracking and surface estimation. In 2009 IEEE Conference on Computer Vision and Pattern Recognition : CVPR 2009, pages 1746–1753, Miami, USA, 2009. IEEE. 1

[8] N. Kasuya, R. Sagawa, R. Furukawa, and H. Kawasaki. Robust and accurate one-shot 3D reconstruction by 2C1P system with wave grid pattern. In Proc. International Conference on 3D Vision, Seattle, USA, June. 2013. 4, 6

[9] 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. 2

[10] M. Kazhdan, M. Bolitho, and H. Hoppe. Poisson surface reconstruction. In Proceedings of the fourth Eurographics symposium on Geometry processing, SGP '06, pages 61–70, 2006. 2, 6, 7

[11] Y. M. Kim, D. Chan, C. Theobalt, and S. Thrun. Design and calibration of a multi-view TOF sensor fusion system. In Computer Vision and Pattern Recognition Workshops, 2008. CVPRW '08. IEEE Computer Society Conference on, pages 1 –7, Jun. 2008. 2

[12] 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

[13] T. Matsuyama, X. Wu, T. Takai, and S. Nobuhara. Real-time 3D shape reconstruction, dynamic 3D mesh deformation, and high fidelity visualization for 3D video. Comput. Vis. Image Underst., 96(3):393–434, Dec. 2004. 1

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

[15] M. Nakazawa, I. Mitsugami, Y. Makihara, H. Nakajima, H. Yamazoe, H. Habe, and Y. Yagi. Dynamic scene reconstruction using asynchronous multiple kinects. In The 7th Int. Workshop on Robust Computer Vision (IWRCV2013), Jan. 2013. 2

[16] 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

[17] 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, 3, 4

[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. 2

[20] D. Vlasic, P. Peers, I. Baran, P. Debevec, J. Popovi, S. Rusinkiewicz, and W. Matusik. Dynamic shape capture using multi-view photometric stereo. ACM Trans. Graphics (Proc. SIGGRAPH Asia), 28(5), 2009. 1, 2

[21] 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