Multiview Projectors/Cameras System for 3D Reconstruction of Dynamic Scenes

Ryo Furukawa, Ryusuke Sagawa, Amael Delaunoy, Hiroshi Kawasaki

Multiview Projectors/Cameras System for 3D Reconstruction of Dynamic Scenes

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

Ryusuke Sagawa
National Institute of Advanced
Industrial Science and Technology,
Tsukuba, Japan
ryusuke.sagawa@aist.go.jp

Amael Delaunoy, Hiroshi Kawasaki
Kagoshima University,
Kagoshima, Japan
a.delaunoy@ibe.kagoshima-u.ac.jp
kawasaki@ibe.kagoshima-u.ac.jp

Abstract

Active vision systems are usually limited to either partial or static scene reconstructions. In this paper, we propose to acquire the entire 3D shape of a dynamic scene. This is performed using a multiple projectors and cameras system, that allows to recover the entire shape of the object within a single scan at each frame. Like previous approaches, a static and simple pattern is used to avoid interferences of multiple patterns projected on the same object. In this paper, we extend the technique to capture a dense entire shape of a moving object with accuracy and high video frame rate. To achieve this, we mainly propose two additional steps; one is checking the consistency between the multiple cameras and projectors, and the other is an algorithm for light sectioning based on a plane parameter optimization. In addition, we also propose efficient noise reduction and mesh generation algorithm which are necessary for practical applications. In the experiments, we show that we can successfully reconstruct dense entire shapes of moving objects. Results are illustrated on real data from a system composed of six projectors and six cameras that was actually built.

1. Introduction

Dense entire shape acquisition of moving objects (*e.g.* human body, bursting balloon, etc.) is strongly required for many fields and has been widely researched so far. To acquire entire shapes, *shape from silhouette* [11] techniques have been commonly used. However, concavity or detailed shapes cannot be recovered with a silhouette based technique. *Multi-view stereo (MVS)* [5, 6, 13] is another solution to reconstruct the entire shape. However, it is still difficult to reconstruct dense and accurate shapes with MVS if there is no texture on the object, or if the object is not perfectly Lambertian. Moreover, if it can recover accurate dense shapes, MVS is usually far from being real-time.

On the other hand, active 3D scanning systems are widely used for practical purposes, because of its accuracy and fidelity. In particular, structured light systems that are composed of a projector and a camera are increasingly developed and produced because high quality video projectors are now available at inexpensive prices. More importantly, scanning time of such systems is really fast and is suitable for dynamic reconstruction. However, there remains several essential problems to use multiple sets of them, surrounding the target object, to capture its entire shape. For example, a commonly used temporal-encoding-based projector-camera system, requires a number of images in which the object is captured with different patterns. Therefore, extending it to multi-projectors is not straightforward, requires a complex synchronization setup and is not applicable for fast motion.

Spatial-encoding-based technique is another solution for projector-camera systems to reconstruct the shape. Since it requires just a single captured image of the object, on which static pattern is projected, no synchronization problem occurs. However, the patterns are usually complicated and they interfere each other if they are projected onto the same object. Recently, the solution for the complex pattern by using a simple grid pattern which embeds information in relation of connection of parallel lines has been published [9, 12, 15]. Furukawa *et al*. extended the method to multi-camera configuration [4]. However, the technique requires a special camera and projector configuration for stable reconstruction. Moreover, independent reconstructions for each set may result in inconsistent shape after integration.

In this paper, we propose a multiple-view reconstruction technique specialized for projector-camera systems based on spatial-encoding techniques. In our approach, mainly two techniques are proposed to solve aforementioned problems; the first one is a consistency check between multiple cameras and projectors and the second one is a plane parameter optimization algorithm for light sectioning methods.

In the experiments, we actually construct the system, which consists of six projectors and six cameras, and successfully reconstruct the series of entire shape of dynamically moving human bodies which has not been succeeded yet in [4]. We also conducted comparisons between state-

of-the-art MVS technique and ours. The main contributions of the paper are as follows:

1. We present a practical system for entire shape acquisition of dynamically moving objects using multiple projectors and cameras.
2. We propose a global shape optimization technique to refine the 3D shape by plane adjustment (Sec. 3.3)
3. We propose an efficient noise reduction method (Sec. 4.1).
4. A mesh generation algorithm from point clouds using Graph Cut and signed distance field is developed (Sec. 4.2)
5. Results are evaluated by comparing to passive MVS.

Related works

For practical 3D acquisition purposes, active sensors have been widely used and a large number of projector-camera-based systems are intensively researched because of its efficiency and simple installation [1, 17]. Projector-camera-based 3D reconstruction approaches can be categorized into two types: temporal and spatial encoding techniques. Although temporal encoding methods can achieve dense and accurate reconstruction, using multiple set of temporal encoding systems is rather difficult, because they simultaneously project multiple set of patterns onto a single object.

Techniques using only a spatial encoding pattern allow scanning with only a single-frame image [8, 16] and are suitable for increasing the number of devices. Since they typically use complex patterns or colors for the decoding process, results usually have ambiguities on textured object or near depth boundaries. Recently, the solution for the complex pattern by using a simple grid pattern which embeds information in relation of connection of parallel lines has been published [9, 12, 15]. However, since the system projects dense vertical and horizontal grid patterns, it is difficult to decompose them after an added projection on the same area. If the pattern is composed of directional parallel lines with one or two colors, those problems can be drastically reduced, see for instance [10]. In [10], the authors assume a single-camera configuration which requires some constraints (*i.e.*, the center of the camera should be on the focal plane of the projectors) in the arrangement of the positions of the camera and the projectors to increase the stability of the solution. Since this is a disadvantage of the method to be extended to multi-camera configuration, they removed the requirement and extended the method to multi-camera configuration [4]. Instead of restricting the devices' configuration, they proposed to use the information of multiple cameras to increase the stability of the solutions. However, the algorithms to achieve that goal were not described in [4], and experiments were only applied to static objects.

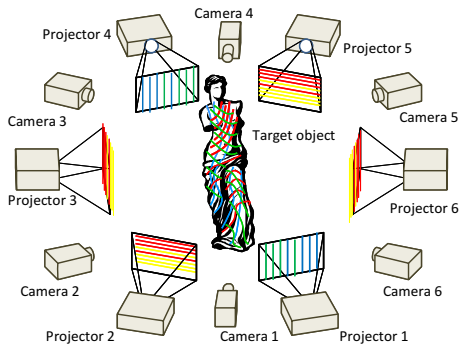


Figure 1. A setup example to reconstruct the entire shape using six projectors and six cameras.

In this work, we extend the work of [4] and detail the algorithms for multi-camera and multi-projector. In particular, we present a cost function to evaluate the quality of the matches by checking the consistency between the different views. In addition, we propose novel post-processing methods to increase the quality of the solution such as noise filtering, correction of errors caused by calibration or pattern detection, and mesh generation from the point cloud results. The efficiency of the method is validated by experiments applied to dynamic scenes of human bodies.

2. Overview

2.1. System configuration

To capture entire shapes of dynamic scenes, we propose a novel 3D measurement system composed of multiple cameras and projectors. Typically, those devices are placed so that they encircle the target scene, and the cameras and the projectors are put alternatively. The projected pattern is static and does not change, thus, *no synchronization is necessary*. A setup example of the system is shown in Fig. 1. All the devices are assumed to be calibrated (*i.e.*, known intrinsic parameters of the devices as well as their relative positions and orientations). The projectors and cameras are placed so that each camera can observe the line patterns projected by adjacent projectors. In the paper, we developed an experimental system with six projectors and six cameras.

To capture scenes, each of the projectors projects parallel single directional lines, and the cameras capture the projected line patterns as 2D curves on the captured images. For each of the detected curves, the projector that generates the pattern is identified by using information of colors and angles of the curves. Then, the intersection points between the detected 2D curves are extracted.

Even though the proposed reconstruction method can reconstruct shapes from monochrome line patterns, in this work we use color patterns to improve accuracy and robustness. In this paper, a periodic pattern of two colors is used based on the de Bruijn sequence [8]. The curve detection is similar to the one described in [4].

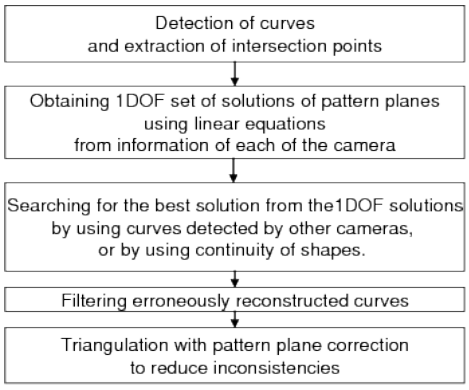


Figure 2. Flow of the proposed algorithm.

2.2. Overview of the reconstruction algorithm

The flow of the reconstruction algorithm is shown in Fig. 2. First, we detect curves with three colors and three directions based on [4] which is an extended version of [12]. Then, we perform reconstruction for each camera by only using single-camera information. This reconstruction can be solved up to 1-DOF indeterminacy and many candidates for the solution remains. This part is described in Sec. 3.1.

Secondly, from those candidates, the optimal solution is selected by using information of adjacent camera pairs. This is done by our multi-view reconstruction algorithm, one of our contributions. By this reconstruction step, the correspondences between the detected curves and the projected line-patterns can be determined, and consequently 3D reconstruction can be achieved. This part is described in Sec. 3.2. In the obtained solutions, erroneously reconstructed curves can be included. Filtering out those curves is applied before global optimization (Sec. 4.1).

Due to errors in the calibration and line-detection, there may still remain inconsistencies in the reconstructed shape. Namely, identical curves detected by multiple cameras are reconstructed at different positions if a number of cameras and projectors are used. In the final step, these gaps are reduced using error correcting triangulation, which is a global optimization based on the consistency of all the cameras and projectors. This is also one of our contribution (Sec. 3.3).

3. 3D shape from multiple projectors/cameras

3.1. Obtaining candidates of the solution

In this section, we summarize the work of [4]. In particular we show how to compute the 1-DOF solution using only single-camera information as in [4]. First, we have curves detected by camera 1 using the curve detection algorithm presented in [4]. We assume that the source projector of each of the detected curves is already identified.

A line pattern projected from a projector goes through a 3-D plane. This plane is called a pattern plane p, represented by an equation $\mathbf{p}^\top \mathbf{x} + 1 = 0$, where $\mathbf{p}$ is the 3D vector that parametrizes the plane, and $\mathbf{x}$ is a point on that plane.

Pattern planes generated by a projector share a single line, which is called an axis of the planes. The variations of the pattern planes sharing a single axis can be defined by a rotation around the axis, thus, it is 1-DOF variation. One of the 1-parameter representation of the planes is $\mathbf{p} = \mathbf{q}\mu + \mathbf{r}$ where μ is the parameter, $\mathbf{q}$ and $\mathbf{r}$ are constant vectors that can be calculated from the positions of the axis.

Assume an intersection between patterns A and B is detected by camera 1, and represented by a vector $\mathbf{u}$ in the camera coordinates of camera 1. Parameter vectors of pattern planes A and B are $\mathbf{p_A}$ and $\mathbf{p_B}$ in the same coordinates. Then, from the works of [12, 15],

$$\mathbf{u}^\top(\mathbf{p_A} - \mathbf{p_B}) = 0. \tag{1}$$

By using 1-parameter representations $\mathbf{p_A} \equiv \mathbf{q_A}\mu_A + \mathbf{r_A}$, and $\mathbf{p_B} \equiv \mathbf{q_B}\mu_B + \mathbf{r_B}$, we obtain:

$$C_A\mu_A - C_B\mu_B + D = 0\,, \tag{2}$$

where $C_A \equiv \mathbf{u}^\top\mathbf{q_A}$, $C_B \equiv \mathbf{u}^\top\mathbf{q_B}$ and $D \equiv \mathbf{u}^\top(\mathbf{r_A} - \mathbf{r_B})$. This is a constraint between plane parameters μ_A and μ_B from a detected intersection.

If N curves are detected by camera 1, they are represented by N parameters, $\mu_1, \mu_2, \cdots, \mu_N$. Let M intersections be observed between the curves, where $M > N$ (this is normally true). From the form in Eq. (2), we obtain M equations with N variables. This can be represented by:

$$\mathbf{C}\,\mathbf{m} = \mathbf{d}\,, \tag{3}$$

where $\mathbf{C}$ is $M \times N$ matrix and $\mathbf{m} \equiv (\mu_1, \mu_2, \cdots, \mu_N)^\top$.

Typically, the solution of Eq. (3) is unstable. However, if one of the N variables are fixed, the other variables can be calculated. This can be explained by the following: If the distribution of the direction vectors of M intersections $\mathbf{u_1}, \mathbf{u_2}, \cdots, \mathbf{u_M}$ is in small area (*i.e.*, $\mathbf{u_1} \approx \mathbf{u_2} \approx \cdots \approx \mathbf{u_M}$), the minimum eigenvalue of $(\mathbf{C}^\top\mathbf{C})^{-1}$ becomes nearly 0, which is almost a degenerate condition. This can happen in a real measurement situation, and it can make the calculation of the linear solution unstable.

On the contrary, Eq. (3) can be stably solved up to the above 1-DOF ambiguity. Intuitively, this is explained by the following: If the parameter of curve A, μ_A, is fixed (*i.e.*, pattern plane of curve A is fixed), the shape of curve A is reconstructed. Using these depth values, pattern planes of the curves that have intersections with curve A can be identified. By repeating these processes, all the connected curves can be calculated. Thus, Eq. (3) is stable except for the 1-DOF ambiguity.

From the above discussion, Eq. (3) can be stably solved up to 1-DOF indeterminacy by $\mathbf{m}(t) = \mathbf{g}\,t + (\mathbf{C}^\top\mathbf{C})^{-1}\mathbf{C}^\top\mathbf{d}$, where $\mathbf{g}$ is the eigenvector of $(\mathbf{C}^\top\mathbf{C})^{-1}$ associated to the minimum eigenvalue, and t is a parameter to represent the variation of solutions. If $\mathbf{C}^\top\mathbf{C}$ is nearly non-regular, $\mathbf{C}\mathbf{m}(t) \approx \mathbf{d}$ for arbitrary t.

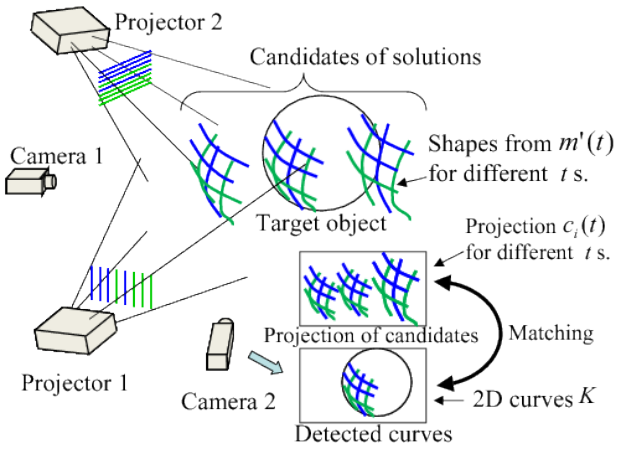


Figure 3. Multi-view reconstruction by two cameras and two projectors: If the 3D curves are at true positions, then, the projected curves should coincide with the curves captured by camera 2.

In the system, the cameras and the projectors are calibrated, and the projected line-patterns are known. Thus, all the candidates of the pattern planes can be calculated and their number is finite. In addition, since we use two colors to encode labels for each line as described in the overview of the algorithm, we can efficiently reduce the 1-DOF ambiguity as follows. Since each of the candidate pattern planes is associated to a certain label of the color code, each element of the set of pattern planes $\mathbf{m}(t)$ for arbitrary t can be corrected to the nearby plane that has the same label of the color code from the finite set of known pattern planes. The corrected set of pattern planes is written as $\mathbf{m}'(t)$.

3.2. Multi-view reconstruction of connected curves

In the previous section, we showed how to obtain the candidates using a single camera with multiple projectors. Since multiple images may see a 3D point, other cameras can be used to increase the stability. The idea was mentioned in [4], but not detailed. In this section, we explain how to account for multi-view information in the multi-camera and multi-projector configuration.

The solutions having the 1-DOF ambiguity described in the previous section is based on the intersections observed by camera 1. By applying triangulation to pattern planes $\mathbf{m}'(t)$ for t, the curves observed by camera 1 is reconstructed. These results are then verified by projecting onto camera 2. This idea is illustrated in Fig. 3. Let $c_i(t)$ be a projection of curve i reconstructed from $\mathbf{m}'(t)$ to camera 2. We define the matching score of $\mathbf{m}'(t)$ by:

$$S_T(t) \equiv \sum_i S_C(c_i(t), K), \tag{4}$$

where $S_C(c, K)$ is a matching score between a 2-D curve c and a set of 2-D curves K, $S_T(t)$ is the total matching score for all the curves reconstructed from $\mathbf{m}'(t)$. The score becomes smaller if the curves match better. See Fig. 3.

From all the candidates $\mathbf{m}'(t)$ for different ts, the solution of best-matching is selected (the middle candidate in Fig. 3). This can be done by selecting $\mathbf{m}'(t^*)$, where $t^* \equiv \arg\min_t S_T(t)$. Then, corresponding pattern planes for each detected curves are decided from $\mathbf{m}'(t^*)$, and the curves are reconstructed.

Note that this method solves the 1-D ambiguity of depth by using matching with other cameras. Thus, this is similar to stereo matching. In the proposed method, search is processed for the 1-DOF indeterminacy, instead of along the epipolar lines as used in usual stereo matching approaches.

A difference between the proposed formulation and the usual stereo methods is that all the curves in a connected set are simultaneously matched (note that each of the candidates in Fig. 3 is a set of curves). In normal stereo matching methods, each point is matched independently. This means that, if a proper matching function is defined, all the curves in a connected set can be correctly reconstructed, even if only a small part of the curves can be matched.

Another advantage of the method is that the proposed approach is more efficient than most multi-view stereo matching methods, since the search space of the solution is 1-DOF for a set of connected curves, which is much less than the n-DOF search space of stereo matching of n points.

Next, we discuss the definition of $S_C(c, K)$. Many of the curves observed by camera 1 cannot be observed by camera 2, because of occlusions or detection failures. Thus, if $S_C(c, K)$ is defined to be a plus value when a corresponding curve of curve c cannot be found in K, the plus values of unobserved curves are accumulated to the score Eq. (4). This leads to incorrect evaluation of the matching.

To deal with this problem, $S_C(c, K)$ is defined as 0 if a corresponding curve of curve c cannot be found in K, otherwise, it is defined such that $S_C(c, K) < 0$, where

$$S_C(c, K) \equiv \sum_{\mathbf{x_c} \in c} \min(0, \min_{k \in K} \min_{\mathbf{x_k} \in k} (\|\mathbf{x_c} - \mathbf{x_k}\| - W)), \tag{5}$$

where $\mathbf{x_c}$ is a point on c, k is a curve in K, $\mathbf{x_k}$ is a point on k, and W is a size of neighborhood from c. Points within distance W from curve c are seen as nearby points to the curve, and affect the score. Points that are not within distance W do not affect the score because of the first $\min$ function.

3.3. Entire shape optimization using inter-curve consistency

For passive stereo reconstruction including multi-view stereo, how to generate consistent 3D shapes from correspondences affected by noise has been a major problem. To achieve this goal in stereo vision, error correcting triangulation methods such as the Hartley-Sturm method [7], or bundle adjustment are widely used.

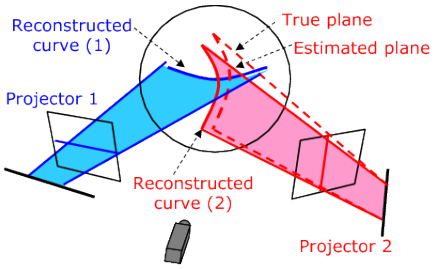


Figure 4. Projector-camera-projector(PCP) case

The triangulation in the proposed system is based on a light-sectioning method, where a 3D point is reconstructed for the intersection point between a line-of-sight and the pattern planes. This is different from stereo vision, where 3D points are reconstructed for intersections between two lines-of-sights. Thus, error correcting triangulation used for passive stereo vision cannot be used in the proposed system.

If naive triangulation of light-sectioning method is applied, following inconsistency occurs:

Projector-camera-projector(PCP) case: Two curves that should intersect at the detected intersection point does not intersect, but there remains some distance between them;

Camera-projector-camera (CPC) case: If a single curve that is observed by multiple camera, the reconstruction results of the multiple observation do not coincide with each other, and it consequently result in different 3D shapes.

An example of the PCP case caused by erroneous position of the pattern plane is shown in Fig. 4. These gaps between curves become a major problem to produce unified shape model from the result. To deal with it, we propose a method to reduce these errors by optimizing the pattern planes so that the above gaps are minimized.

Let p_i be each of the reconstructed pattern planes. We assume p_i can be corrected by rotating it around the axis of the plane (the axis is defined from the source projector) by an infinitesimal angle θ_i. Using the correction, the errors of the above cases of PCP and CPC are minimized.

For PCP cases, suppose that pattern planes p_i and p_j (in Fig. 4, i and j correspond to 1 and 2, respectively) share an intersection point I, and p_i and p_j are corrected by angles θ_i and θ_j, respectively. Let $f_i(\theta_i)$ be the depth from the observing camera to the reconstructed point I calculated by triangulation with plane p_i, and $f_j(\theta_i)$ be the depth of I calculated by triangulation with p_j. Then, the gap between the two reconstructed curves can be defined as $f_i(\theta_i) - f_j(\theta_j)$.

For CPC cases, let c_k be a curve observed by camera C_i, and c_l be one observed by camera C_j. Suppose that both c_k and c_l corresponds to a single pattern plane p_m projected from projector P. Then, c_k and c_l may share physically same curves. To detect the shared points, a distance between 2D curves are defined, and nearby curve c_l of camera C_j from curve c_k is searched, where curve c_k is and a 2D projection into camera C_j of a 3D curve reconstructed using pattern plane p_i. Let points p_s and p_t be nearby points that are sampled from curves c_l and c_k, respectively. If the distance is less than a threshold, p_s and p_t are regarded as the same point, and the pattern planes are corrected such that distances between those pair of points becomes small. Let $g_s(\theta_m)$ and $g_t(\theta_m)$ be the depths from projector P of the points p_s and p_t reconstructed by pattern plane p_m corrected by angle θ_m, respectively. Then, the error of the case CPC for this pair of points can be represented by $g_s(\theta_m) - g_t(\theta_m)$. The sum of two types of errors described above can be defined as:

$$E(\theta_1, \theta_2, \cdots) = \sum_{p \in P} \theta_p^2 + W_c \sum_{(i,j) \in C} (f_i(\theta_i) - f_j(\theta_j))^2 + W_s \sum_{(s,t,m) \in S} (g_s(\theta_m) - g_t(\theta_m))^2 . \quad (6)$$

where W_c and W_s are weights to control significances of the constraints, P is the set of reconstructed patterns, C is the set of detected intersection points, and S is a set of pairs of points that are regarded as the same points and the projector of the corresponding patterns.

The error E of form (6) is minimized under the assumption: $\theta_1 \approx \theta_2 \approx \cdots \approx 0$. This can be approximately solved as the least square solution of the following linear equations:

$$\theta_p = 0 \ \ (p \in P) , \quad (7)$$

$$\frac{\partial f_i}{\partial \theta_i}\theta_i - \frac{\partial f_j}{\partial \theta_j}\theta_j = -f_i(0) + f_j(0) \ \ ((i,j) \in C) , \quad (8)$$

$$\frac{\partial g_s}{\partial \theta_m}\theta_m - \frac{\partial g_t}{\partial \theta_m}\theta_m = -g_s(0) + g_t(0) \ \ ((s,t,m) \in S). \quad (9)$$

The coefficient value represented as the partial derivatives can be either calculated analytically or by approximation using numerical differences. By reconstructing the curves from the pattern planes corrected by the obtained solution, the errors can be reduced.

4. Shape refinement algorithm

4.1. Filtering erroneously reconstructed curves

Using the method described in the previous section, correspondences between the detected curves and the patterns can be determined. These correspondences may include errors, such as wrong connection between different curves, or detection errors of color codes.

For example, since these curves may affect the entire shape optimization described in Sec. 3.3, they should be removed before the process. As shown in Fig. 5. At a detected intersection on the incorrect curve (*i.e.*, curve (1) in Fig. 5), there are 3D gaps between curve (1) and curve (2).

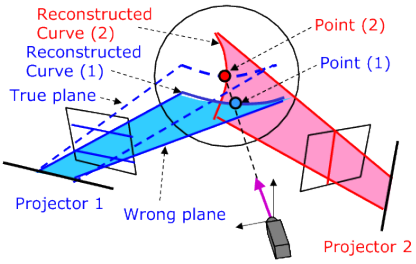
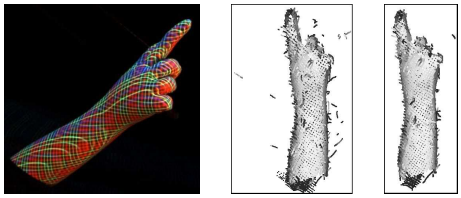


Figure 5. Filtering algorithm



(a) (b) (c)

Figure 6. An example of filtering: (a) source image, (b) the result without filtering, (c) the result after filtering.

To filter out such curves, the intersection points used for reconstruction are re-checked for the reconstructed curves. If the distance between the 3D points (points (1) and (2) in Fig. 5)), which are reconstructed from intersection on 2D image, is larger than a tolerance threshold, the points are labeled as invalid. Curves that have many invalid intersection points are rejected from the reconstruction.

The effect of such filtering is shown in Fig. 6. The number of isolated curves is reduced by this filtering step.

Depending on conditions of the image capturing, extraction of silhouette data from input images is possible. In this case, further noise reduction is possible by projecting each of the reconstructed points to the silhouette images and removing points that are projected outside the silhouettes.

4.2. Polygon mesh generation

From the reconstruction algorithm described in Sec. 3, the 3D positions of the observed curves are obtained. Although these curves are sparse in the images, per pixel distance images can be obtained using the following algorithm.

In the process of curve detection, we analyze the adjacency information between curves. For the region occupied by a set of curves that are related by adjacency, we apply 1D Gabor filter. Then, per-pixel information of 'local' 1D projector coordinates that are defined relatively in the curve set can be calculated in sub-pixel accuracy.

In addition, for the curves that are reconstructed by the proposed method, the 'global' 1D projector coordinates are absolutely defined. By using the local coordinates with global coordinates of the neighbor pixels where curves are detected, global coordinates of each pixel can be calculated. Since there may be multiple pixels having global coordinates in the neighbor of one pixel, the global coordinates can be noise-filtered by using average or median values. Per-pixel depth images are immediately obtained from the projector coordinates.

By integrating the dense depth images, a polygon mesh surface can be obtained. In this study, we implemented a polygon-mesh generation method by Graph Cut using volumetric signed distance fields, which is an extention of [3].

5. Experiments

5.1. Evaluation by synthetic data

To confirm the effectiveness of the proposed method, we first use synthetic data for reconstruction. We use the Stanford bunny as the target. We virtually set six cameras and six projectors surrounding the target object as shown in Fig. 7 (a) and render the captured images by using POV-Ray as shown in Fig. 7 (b).

By using the intersection of detected curves of multiple projectors, polygon mesh is reconstructed using the proposed multi-view reconstruction and the mesh generation technique as shown in Fig. 7 (c)-(d). We also reconstruct the shape with state-of-the-art MVS [2] and previous method [4] as shown in Fig. 7 (e)-(h). For the passive MVS [2], since projectors are also used for projecting patterns onto the object to realize accurate reconstruction with only six cameras, we considered that it is appropriate to use the same setup for comparison (not using 12 cameras).

From the figures, we can see that most of the parts are successfully recovered with all the methods. Since the number of the camera is just six, several parts, especially for occluded areas, are wrongly reconstructed with MVS. Comparing to the previous method, there remain many noises by wrong curve detection and large holes for occluded areas, whereas those problems are solved in our method. RMSE of reconstructed shapes are 0.0023, 0.0112 and 0.0089 for our method, MVS and previous method [4], respectively (the height of the bunny was scaled to be 0.1).

5.2. Entire shape acquisition with a real system

Next, we reconstruct the entire shape of the object with the actual system (See Fig. 8). We use six cameras of Point Grey Research Grasshopper (1600×1200 pixels), and six projectors of LCD video projectors with XGA resolution. The cameras are synchronized and can capture images at 30 frames/sec. The devices are pre-calibrated before the experiments. The intrinsic parameters of cameras and projectors are hard-calibrated by using OpenCV and the extrinsic parameters of the system are self-calibrated by bundle adjustment which is provided by Snavely *et al*. [14]. The corresponding points between a camera and a projector are given by projecting time-multiplexed structured-light patterns from a projector.

3D reconstruction results with our technique and the MVS [2] are shown in Fig. 9(a)-(h). From the figures, we can see that the human body are successfully recovered with both techniques when the persons' pose is simple, whereas

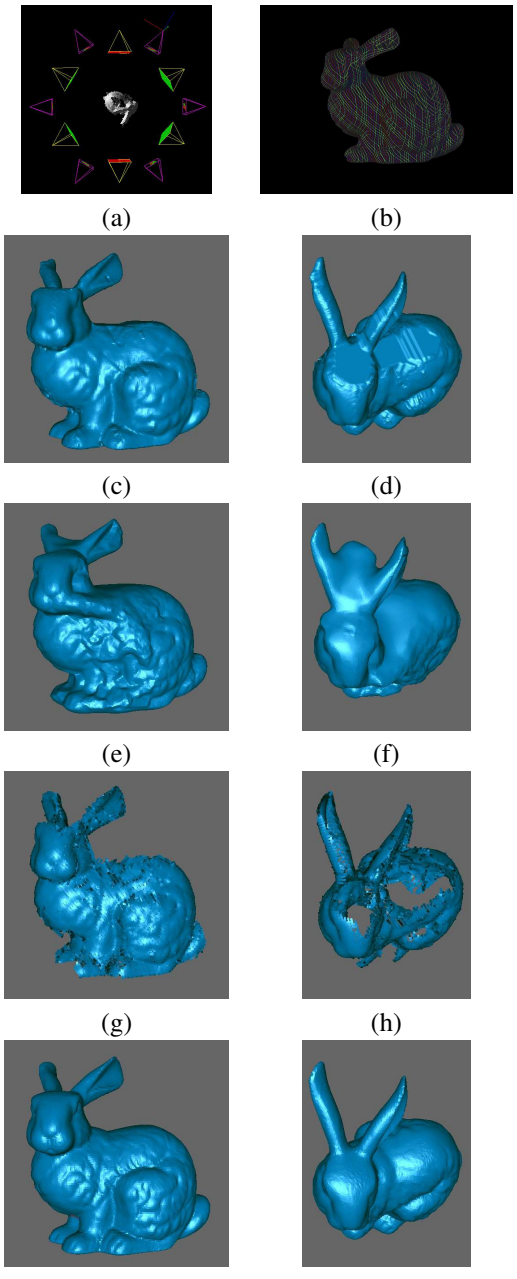


Figure 7. Reconstruction results: (a) configuration of cameras and projectors, (b) synthesized image, (c)-(d) recovered shape with our technique, (e)(f) recovered shape with MVS [2], (g)(h) recovered shape by Furukawa *et al*. [4], and (i)(j) ground truth.

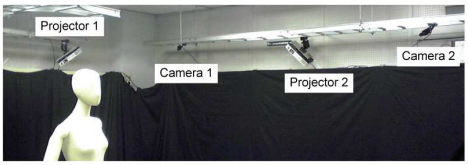


Figure 8. An experimental setup composed of 6 cameras and 6 projectors.

several artifacts are found with both techniques when the persons' pose is complicated. With MVS, since texture information are efficiently used, detailed shapes are successfully reconstructed. With our technique, occluded areas that are seen only by one camera are correctly recovered. In addition, our algorithm runs in a few seconds, where MVS takes around 20 minutes. While combining the two complementary approaches seems promising, this is out the scope of this paper and is part of future work.

5.3. Entire shape acquisition of dynamic objects

Finally, we show the results by capturing moving objects at high frame rate. First, we captured a scene where a person plays Judo with Kimono. Fig. 10(a) shows the images of a frame captured by the cameras. Fig. 10(b) is the reconstructed result rendered from various view points. Next, we captured a scene where a woman is dancing with skirt. Fig. 10 (c)shows the input images and the reconstructed results are shown in Fig. 10(d). The entire shape of the person is reconstructed as shown in the figures. Although some parts are missing due to the occlusion, the proposed method successfully reconstructed the 3D shapes of persons. The computational time by a PC with Intel Core i7 2.8GHz was 726 seconds for the first sequence (30 frames). The average times for a frame were 13.5 and 10.7 seconds for image processing and 3D reconstruction, respectively. It can be reduced by introducing parallel and GPU computation for each camera and this is our future development target.

6. Conclusion

In this paper, a multi-view projectors and cameras system for capturing an entire shape of a moving object was presented. The 3D reconstruction of our algorithm is achieved by first obtaining 1-DOF solution using information from a single camera similarly as in [4], and then, searching the optimum solution from the candidate solutions using consistency between the adjacent cameras. We also proposed error-correcting triangulation that reduces gaps between reconstructed curves. It was achieved by simultaneously minimizing the gaps for the entire shape. In our experiment, we constructed a system that consists of six projectors and six cameras and captured moving objects successfully.

Acknowledgment

This work was supported in part by SCOPE No.101710002, Grant-in-Aid for Scientific Research No.21200002 and NEXT program No.LR030 in Japan.

References

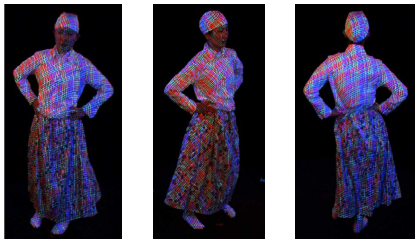
[1] J. Batlle, E. M. Mouaddib, and J. Salvi. Recent progress in coded structured light as a technique to solve the correspondence problem: a survey. *Pattern Recognition*, 31(7):963–982, 1998.

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

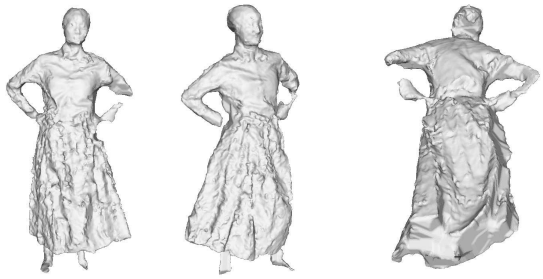
Figure 9. Result of multi-view reconstruction: (a) (c) (e): result of MVS [2] and (b) (d) (f): result of our technique.

(a) Input of Judo sequence.

(b) Result of Judo sequence (without mesh generation).



(c) Input of dancing sequence.



(d) Result of dancing sequence (with mesh generation).

Figure 10. Experiments of dynamic scenes: (a) the input images captured by 6 cameras, (b) the reconstructed results, (c) the input images of dancing sequence, and (d) reconstructed results.

[2] A. Delaunoy and E. Prados. Gradient flows for optimizing triangular mesh-based surfaces: Applications to 3D reconstruction problems dealing with visibility. *IJCV*, pages 1–24, 2010.

[3] R. Furukawa and H. Kawasaki. Shape-merging and interpolation using class estimation for unseen voxels with a gpu-based efficient implementation. In *IEEE Conf. 3DIM*, pages –, 2007.

[4] 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.

[5] Y. Furukawa and J. Ponce. Accurate, dense, and robust multi-view stereopsis. In *CVPR*, 2007.

[6] Y. Furukawa and J. Ponce. Dense 3D motion capture from synchronized video streams. In *CVPR*, 2008.

[7] R. Hartley and P. Sturm. Triangulation. *CVIU*, 68(2):146–157, 1997.

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

[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.

[10] H. Kawasaki, R. Furukawa, R. Sagawa, Y. Ohta, K. Sakashita, R. Zushi, Y. Yagi, and N. Asada. Linear solution for oneshot active 3d reconstruction using two projectors. In *3DPVT*, 2010.

[11] A. Laurentini. How far 3d shapes can be understood from 2d silhouettes. *IEEE Trans. on PAMI*, 17(2):188–195, 1995.

[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.

[13] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *CVPR*, volume 1, pages 519–528, 2006.

[14] N. Snavely, S. M. Seitz, and R. Szeliski. Photo tourism: Exploring image collections in 3D. In *ACM SIGGRAPH*, 2006.

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

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

[17] M. Young, E. Beeson, J. Davis, S. Rusinkiewicz, and R. Ramamoorthi. Viewpoint-coded structured light. In *CVPR*, June 2007.