

A Triangle Mesh Reconstruction Method Taking into Account Silhouette Images

Michihiro Mikamo¹, Yoshinori Oki¹, Marco Visentini-Scarzanella¹,
Hiroshi Kawasaki¹, Ryo Furukawa², and Ryusuke Sagawa³

¹ Graduate School of Science and Engineering, Kagoshima University,
1-21-40, Korimoto, Kagoshima 890-0065 Japan

{mikamo,oki,marco,kawasaki}@ibe.kagoshima-u.ac.jp
<http://www.ibe.kagoshima-u.ac.jp/~cgv/index.html>

² Graduate School of Information Sciences, Hiroshima City University,
3-4-1, Otsukahigashi, Asaminami-ku, Hiroshima 731-3194 Japan

ryo-f@hiroshima-cu.ac.jp
<https://www.hiroshima-cu.ac.jp/>

³ National Institute of Advanced Industrial Science and Technology,
1-3-1 Kasumigaseki, Chiyoda-ku, Tokyo 100-8921 Japan

ryusuke.sagawa@aist.go.jp
https://www.aist.go.jp/index_en.html

Abstract. In this paper, we propose a novel approach to reconstruct triangle meshes from point sets by taking the silhouette of the target object into consideration. Recently, many approaches have been proposed for complete 3D reconstruction of moving objects. For example, motion capture techniques are used to acquire 3D data of human motion. However, it needs to attach markers onto the joints, which results in limiting the capturing environments and the number of data that can be acquired. In contrast, to obtain dense data of 3D object, multi-view stereo scanning system is one of the powerful methods. It utilize images taken by several directions and enables to reconstruct 3D dense point sets by using Epipolar geometry. However, it is still challenging problem to reconstruct 3D triangle mesh from the 3D point sets due to the abundant points originated by mismatched points between images. We propose a novel approach to obtain more accurate triangle mesh reconstruction method than the previous one. We take advantage of silhouette images acquired in the process of reconstructing 3D point sets that result in removing noises and compensating holes. Finally, we demonstrate that the proposed method can generate the details of the surface, where the previous method loses from a small number of points.

Keywords: active measurement system, projector-camera system, entire 3D shape, multi-view image reconstruction

1 Introduction

In this paper, we propose a novel approach to generate triangle meshes based on the silhouette of the target object. Recently, 3D data is becoming one of

the most important information to represent the object motion and shape, such as humans, animals and so on. To acquire such 3D data, several techniques have been invented. One such example is motion capture. To detect the motion of objects, markers are attached onto joints. However, the number of samples acquired is limited to the number of markers. As a result, it is difficult to reconstruct surface information using such devices. On the other hand, to obtain dense 3D point sets, various techniques have been proposed in the literature, including Shape-from-Silhouette [14], Multi-view Stereo (MVS) [8]. Shape from Silhouette is one of the typical methods to acquire the entire shape in dynamic scenes, however surface details usually cannot be recovered correctly. Multi-view stereo can generally yield accurate reconstructions. However, reconstructing accurate point sets and generating 3D triangle mesh from the point sets are still challenging tasks.

Multi-view stereo reconstructs 3D point sets by computing corresponding points by the images captured from several directions. The point sets are refined by silhouette images that are the binary images contouring the object, to remove points originated from calibration error or background textures. After reconstructing the point sets, the points are connected to be a triangle mesh. There are some methods that can generate a 3D mesh, however, they tend to fail to connect the points, especially in the case that the number of the points are small, and lose the details of the surfaces.

The main contribution of this paper is that we propose a novel approach that can generate a triangle mesh with a higher accuracy than the state of the art. This is achieved by integrating the silhouette images of the target object into the mesh generating process. The silhouette images prevent abundant points to be connected to generate a 3D mesh. The proposed method is evaluated using point sets acquired by a multi-view projectors/cameras system. We compare the proposed method with a technique representative of the state of the art, which shows our method can generate surfaces from small number of points. Our method can cover holes that tend to appear in sparse point areas and keep the details on the surfaces.

The paper consists of the following sections. In section 2, we briefly introduce some representative methods whose purpose is reconstructing 3D points or meshes. In section 3, we mention the overview of the multi-view projectors/cameras system that enables to reconstruct the point set from images captured by several directions and how to exploit silhouettes from the images. We explain the details of the proposed method in section 4. Experimental results are mentioned in section 5. Finally we conclude the paper by directing future vision in section 6.

2 Related Work

Lots of techniques have been developed to reconstruct 3D shape by using multi-view stereo technique. We review some of representative methods here.

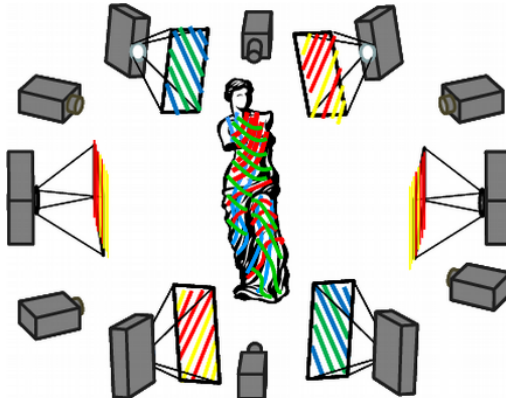


Fig. 1. A setup example to reconstruct the entire shape using six projectors and six cameras.

Several methods use the concept of visual hull such as [5], [9], [14], [16]. Visual hull is computed by taking the intersection of slices of space projected from the target object in the input image. Several techniques utilize visual hull as the initial shape of the 3D model. However, visual hull cannot reconstruct the details of the surfaces. Therefore, improvements are proposed by the articles [5], [9].

Volumetric multi-view stereovision techniques decompose the domain into subdivided areas [2], [3], [6]. These methods tend to be time-consuming or fail to reconstruct 3D shapes because of initial settings of the optimization.

Other approaches for reconstructing 3D shape are well-summarized in Seitz *et al.* [19]. Labatut *et al.* also introduces such techniques by categorizing them from several aspects [13].

Surface reconstruction has been actively developed in computer graphics fields such as [1], [4], [12], and [11]. Those methods are implicitly assuming that the point sets are dense enough to reconstruct 3D surface. However, point sets acquired by multi-view stereo system are not always sufficient. In addition, density of the points would be different depending on the areas. This is because occlusions occur behind the target object, where the reconstruction would be failed.

The proposed method enables to reconstruct surface from relatively small number of points and space areas. In addition, our method can preserve the shape of the surface without smoothing the details.

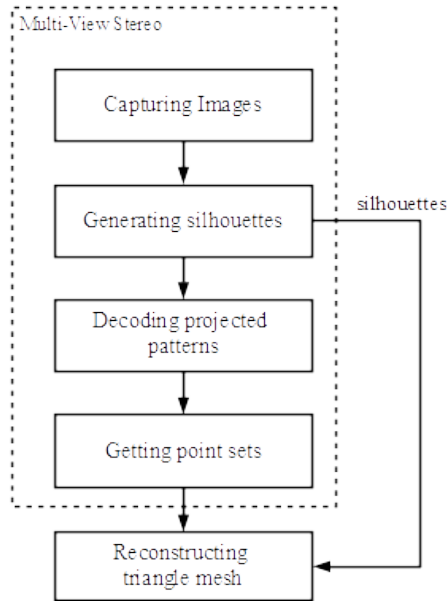


Fig. 2. The overview of the proposed method

3 Overview of the multi-view projector camera system

3.1 System configuration

For our proposed method, we use the active MVS system proposed by Furukawa *et al.* [7], where multiple cameras and projectors are used.

In the setup, devices are placed so that they encircle the target object, and the cameras and the projectors are put in alternating order with known position and orientation. The system was setup as shown in Fig 1. We assume all the devices are calibrated (*i.e.* known intrinsic parameters of the devices as well as their relative positions and orientations). The details of the system configuration are explained in the paper [7].

3.2 The multi-view projector camera system

The system obtains triangle meshes by the following steps; capturing images, generating silhouettes, decoding projected patterns, getting point sets of the target object, and finally, reconstructing triangle meshes. The overview of the proposed method is shown in Fig. 2. The process from capturing images to getting point sets are included in the multi-view stereo system. We use the silhouettes of the target object to generate 3D triangle mesh with higher accuracy.

In order to capture the geometry of the object, each of the projectors projects patterns, while the cameras capture the projected patterns as 2D curves on the captured images.

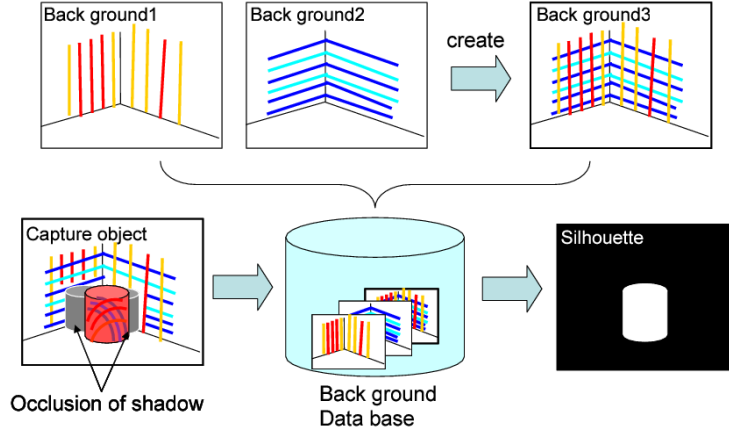


Fig. 3. Workflow for the silhouette extraction.

Next, we generate silhouettes to suppress noise from point mismatches resulting in 3D points outside the target object. To this end, we employ a silhouette generating method based on an offline image database composed by the background images taken under all combinations of projectors, in order to be robust to cast shadows [18]. The details of the silhouette extracting method is explained in Section 3.3. In the proposed mesh generating process, this silhouette image is also used as a term of a cost function, which brings more accurate reconstruction result than the previous method. Wrong matches inside the object don't affect the cost function because they don't appear on surfaces. The tetrahedra composed of the points are also invisible from outside, therefore, they don't affect the final result.

Decoding the projected patterns is necessary to get the shape information of the target object. However, the color information projected is often lost or corrupted due to the texture of the target object. Several methods are proposed to recover the projected patterns despite the interference of the surface texture, such as [21] which we use in this work.

Next, we compute the 3D points by decoding the projected patterns. Namely, we reconstruct the point sets by matching corresponding points between the captured images. Since we use the multi-view system, we can get six sets of points that come from six projectors.

Finally, we reconstruct the triangle mesh representation of the target object. To do so, first, we must find the optimal position for the point sets. We propose a method to determine this position by using the silhouette as a term of the cost function in Section 4.

3.3 Creating silhouette based on the database

In our system, we use the technique described in [18]. The purpose of defining the 3D silhouette of the object is to avoid corresponding points being generated outside of the target object. In addition, we use the silhouette as a term of cost function to prevent that meshes are appeared by misaligned points.

The basic idea of the technique is to create all possible the image areas affected by cast shadows as shown in Fig. 3. Therefore, we capture multiple images of the background illuminated by different subsets of the projectors. All these images are put in a database, which is used to calculate the similarity with each region of the target image. Therefore it allows to find correspondences of background regions even the captured image includes shadows.

During the shape acquisition procedure, similarity is calculated for each patch of the object image against the database using ZNCC (Zero-mean Normalized Cross-Correlation), which allows for linear changes in luminance. However, whenever a candidate patch contains both a subregion with cast shadows as well as one with a clean background, the similarity score is negatively affected. Similarly, at the boundary of the projected lines the perceived color is affected by the demosaicing algorithm, creating instabilities and false matches. In order to alleviate these issues, we consider directly using Bayer pattern images as the input and use an adaptive support-weight [22] when calculating the similarity. Adaptive support-weights work by setting a weight to each pixel of the patch based on color similarity between the center and outer pixels, which we have found experimentally to solve our problem. In all our experiments, we use a 5×5 patch size. The similarity score is used as the data term for a Graph Cut algorithm, which we use for the final segmentation into foreground and background.

4 Triangle mesh reconstruction

In this section, we propose a method that enables to generate triangle meshes from point sets. The purpose of the proposed method is to generate accurate polygon meshes by taking advantage of the information obtained from the multi-view stereo system.

Our method stems from the mesh generating technique proposed in the paper [13]. This consists of constructing a triangle mesh by viewing the problem as binary labeling, where labels stand for the points being either inside or outside of the target shape generated by 3D Delaunay triangulation. For optimization, graph cut was used to minimize the following cost function.

$$E(S) = E_{vis}(S) + \lambda_{photo}E_{photo}(S) + \lambda_{area}E_{area}(S) \quad (1)$$

where, S is a set of triangles belonging to the target mesh. $E_{vis}(S)$ is the visibility that indicates whether the triangle set can be seen from the camera or not. The idea is that the triangle set S consists of the surface of the target object is visible from the camera. $E_{vis}(S)$ is represented by the total number of tetrahedra that

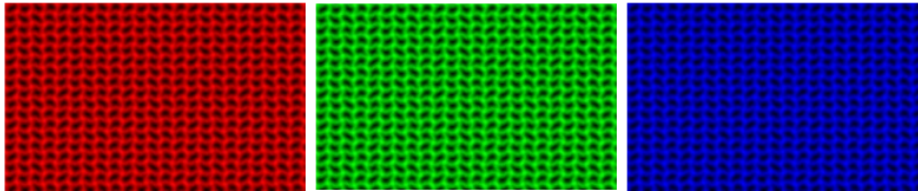


Fig. 4. The wave projection pattern

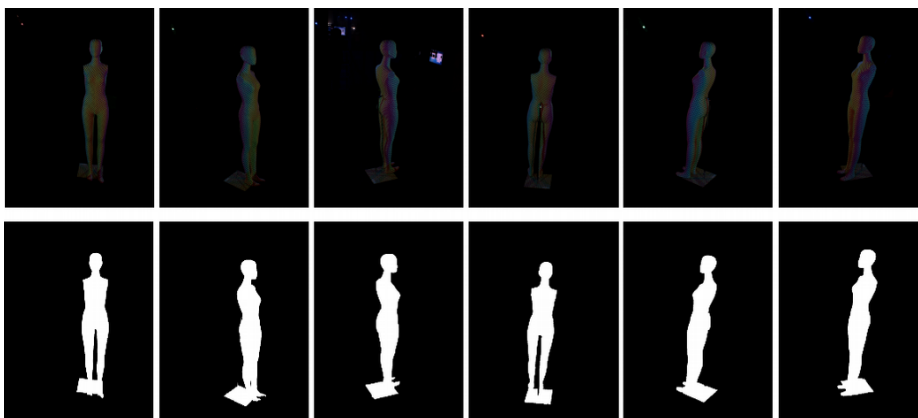


Fig. 5. The captured images (top row) and obtained silhouettes (bottom row)

the ray passes through from each point of a tetrahedron while it reaches the camera. $E_{area}(S)$ is one of the areas of triangles on a tetrahedron.

In our method, we accept the active reconstruction method using multi-view stereo system proposed in the paper [10] to get the point sets of a target object. We used the projection patterns shown in Fig. 4. The advantage of this projection patterns is that we can acquire point sets independent from other projected patterns. The arrangements of the devices are same to Fig. 1.

The process of the proposed triangle-mesh generating method is as follows: First, we merge the point sets to reduce the errors that come from calibration process. Next, we construct 3D Delaunay triangulation by using graph cut. Finally, we solve the optimization problem by Levenberg-Marquadt method to get the resultant triangle mesh.

5 Experiment

5.1 Merging point sets

Our system consists of six cameras and six projectors aligned alternately (See Fig. 1). Therefore, we reconstruct three dimensional points using the combina-

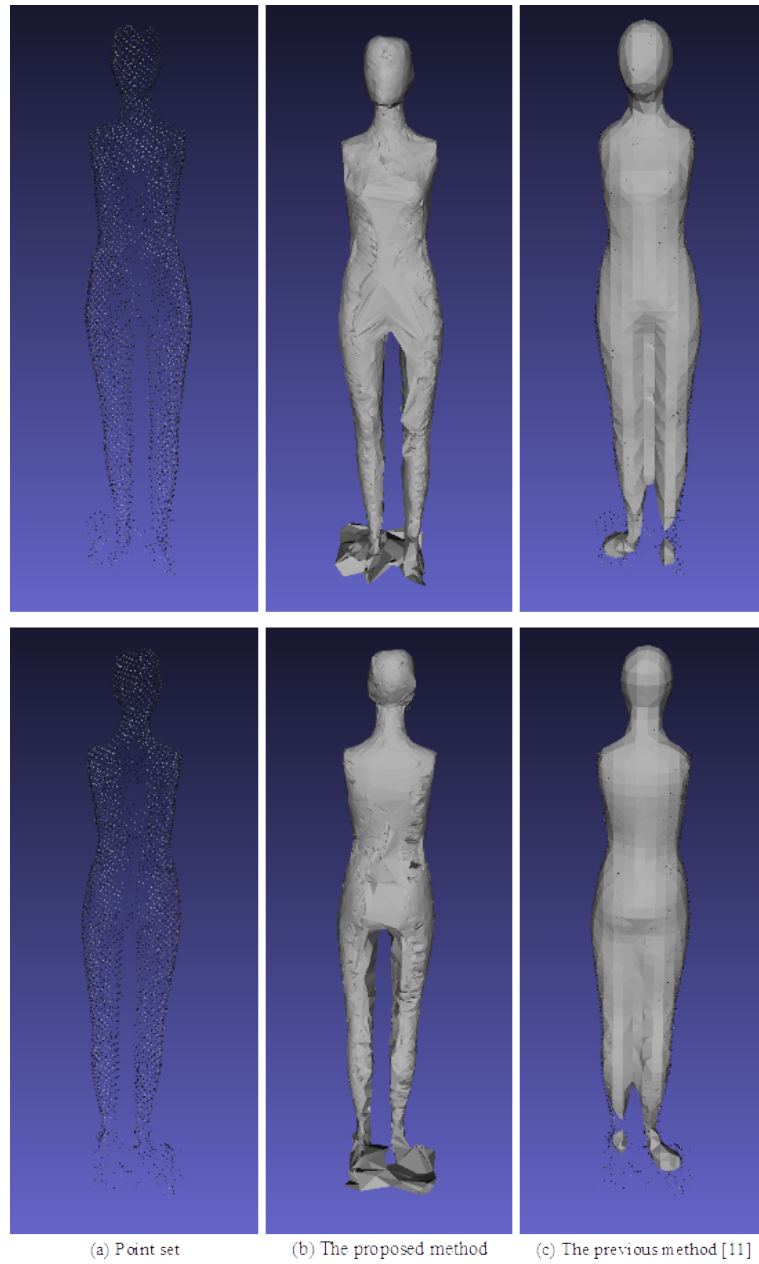


Fig. 6. The comparison between the proposed method (center column(b)) and the previous method [11](right column(c))

tions of cameras and projectors. The minimum component is that a projector is aligned between the two paired of cameras. As a result, each projector has two reconstructed point sets. Then, twelve point sets can be obtained. Point sets obtained by the same projector are supposed to be matched, however, mismatches exist because of the calibration error. We optimize the point sets using SBA (Sparse Bundle Adjustment) [15].

5.2 Computing the Delaunay triangulation in 3D space

In the two dimensional space, Delaunay triangulation connects points to be triangles. In the three dimensional space, it connects points with neighboring three points so that they make a tetrahedron. The circumcircles of the triangle do not include the other points, therefore, it is possible to construct mesh structure by the neighboring points. We construct 3D Delaunay map from 3D point sets. To obtain the 3D Delaunay map, we used The Computational Geometry Algorithms Library (CGAL) [20].

5.3 Graph cut optimization

The method proposed in the paper [13] generates triangle mesh by regarding the problem as a binary labeling. They assign labels to the tetrahedron so that they stand for inside or outside of the target shape. In the proposed method, we follow this concept, however, we propose a new cost function to prevent noise that would be reconstructed from abundant 3D points.

To add the binary labels, we use graph cut optimization. We optimize the following cost function.

$$E(S) = E_{vis}(S) + \lambda_{sil}E_{sil}(S) + \lambda_{area}E_{area}(S) + \lambda_{length}E_{length}(S) \quad (2)$$

λ_{sil} , λ_{area} , and λ_{length} are weights for each term. $E_{sil}(S)$ is the total number of white pixels that appear by a tetrahedra being reprojected to be a silhouette image. By using this term, we generate meshes based on the silhouette images, which result in decreasing noises. $E_{area}(S)$ is the energy that decides the surface smoothness. $E_{length}(S)$ is the energy that stands for the sum of the length of the edges by the triangle that consists of tetrahedra.

5.4 Experimental settings

In our experiment, we used six Point Grey Research Grasshopper cameras (1600×1200 pixels), and six LCD video projectors at WXGA resolution. The cameras are synchronized and calibrated prior to the experiment and can capture images at 20 fps. We used a mannequin model as an object.

5.5 Experimental results

The images captured by the MVS system are shown in the top row of Fig. 5. The wave pattern was projected onto the object. We used the method proposed in the paper [18] to generate silhouette images. The acquired silhouettes images are shown in the bottom row of Fig. 5. The silhouettes are properly obtained without including the texture of background.

Fig. 6 illustrates the comparison result. The top row shows the front view and the bottom row shows the back view. We show the merged point set obtained by our system in Fig. 6(a). The number of vertices is 4,654. Despite the low number of vertices, we are able to generate a smooth polygon mesh as shown in Fig. 6(b). Fig. 6(c) is the result obtained by the method proposed in [11] and implemented on Meshlab [17]. We tried to find the best parameter settings in the method [11], however we found empirically that the suggested parameters were the best: Octree Depth is 6, Solver Divide is 6, Samples per Node is 1, and Surface offsetting is 1. The computation time is about 0.5 [sec.] on the software.

In Fig. 6(b), we can see that the proposed method enables to reconstruct areas where detail tends to be lost, such as the area where the points are occluded by the legs, where the points are sparsely distributed. Conversely, these points are connected into contiguous surfaces when using the method in [11] as shown in Fig. 6(c). In addition, noise that existed in the point set was removed by the proposed method in 6(b), while they can be seen around the object in Fig. 6(c).

Our method still left some bumps. This is caused by the area where the point sets overlapped due to slight calibration errors.

We conducted our experiment under the condition that Intel(R) Core(TM) i7-2600 CPU (3.40GHz), 16GB RAM, x64 system. The program was executed on a single core. The total computation time is 756.31 [sec.]. The most time-consuming part is the collision judgment between the Delaunay diagram and silhouette images. It takes 379.69 [sec.] and accounts for about 50 % of the total computation time. The visibility judgment that the rays count the number of surface across the path takes 362.45 [sec.] and accounts for about 48%.

In Fig. 7, we applied our method to data obtained from a posing human subject. The number of vertices is 16,287. Fig. 7(d) shows some of the captured images. Again, the results clearly show how our method can better preserve the details in scenes with a low number of vertices.

6 Conclusion

We proposed a novel approach to generate triangle meshes by integrating information from the silhouette images of the target object. By using the silhouette images as a term of the cost function, we are able to successfully remove mismatched points that would result in failure cases during the mesh reconstruction. Consequently, the proposed method can generate triangle meshes from relatively small number of points while preserving the surface details. Especially when compared with the state of the art, our method is able to better preserve the surface details with a small number of vertices.

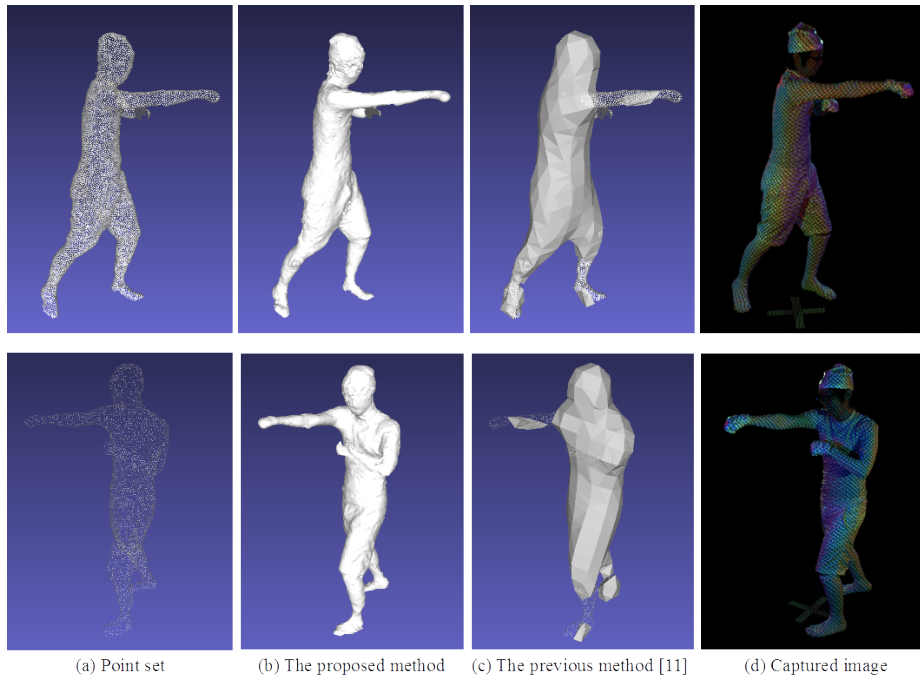


Fig. 7. The resultant mesh obtained from complex object

For our future work, we will explore the parameter space of our proposed method for a better optimization performance. In addition, efficient implementation is another future work. In the proposed algorithm, all the processes can be separated by the minimum projector-camera components. Therefore, the computation time can be reduced by parallel processing such as using GPU.

References

1. J.-D. Boissonnat. Geometric structures for three-dimensional shape representation. *ACM Transactions on Graphics*, 3(4):266–286, 1984.
2. Y. Boykov and V. Lempitsky. From photohulls to photoflux optimization. In *Proceedings of British Machine Vision Conference*, volume 3, pages 1149–1158, 2006.
3. A. Broadhurst, T. W. Drummond, and R. Cipolla. A probabilistic framework for space carving. In *Proceedings of IEEE International Conference on Computer Vision*, pages 388–393, 2001.
4. H. Edelsbrunner and E. P. Mücke. Three-dimensional alpha shapes. *ACM Transactions on Graphics*, 13(1):43–72, 1994.
5. C. H. Esteban and F. Schmitt. Silhouette and stereo fusion for 3d object modeling. *Computer Vision Image Understanding*, 96(3):367–392, 2004.

6. O. Faugeras and R. Keriven. Variational principles, surface evolution, pdes, level set methods, and the stereo problem. *Transactions on Image Processing*, 7(3):336–344, 1998.
7. R. Furukawa, R. Sagawa, A. Delaunoy, and H. Kawasaki. Multiview projectors/cameras system for 3d reconstruction of dynamic scenes. In *Proceedings of 4DMOD Workshop on Dynamic Shape Capture and Analysis*, pages 1602–1609, 2011.
8. Y. Furukawa and J. Ponce. Accurate, dense, and robust multi-view stereopsis. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2007*, pages 1362–1376, 2007.
9. Y. Furukawa and J. Ponce. Carved visual hulls for image-based modeling. *International Journal of Computer Vision*, 81(1):53–67, 2009.
10. N. Kasuya, R. Sagawa, R. Furukawa, and H. Kawasaki. One-shot entire shape scanning by utilizing multiple projector-camera constraints of grid patterns. In *Proceedings of IEEE International Conference on Computer Vision Workshops 2013*, pages 299–306, 2013.
11. 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.
12. R. Kolluri, J. R. Shewchuk, and J. F. O'Brien. Spectral surface reconstruction from noisy point clouds. In *Proceedings of Eurographics/ACM SIGGRAPH Symposium on Geometry Processing 2004*, pages 11–21, 2004.
13. P. Labatut, J.-P. Pons, and R. Keriven. Efficient multi-view reconstruction of large-scale scenes using interest points, delaunay triangulation and graph cuts. In *Proceedings of IEEE 11th international conference*, pages 1–8, 2007.
14. A. Laurentini. The visual hull concept for silhouette-based image understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(2):150–162, 1994.
15. M. I. A. Lourakis and A. A. Argyros. Sba: A software package for generic sparse bundle adjustment. *ACM Transactions on Mathematical Software*, 36(1):1–2, 2009.
16. W. Matusik, C. Buehler, R. Raskar, S. J. Gortler, and L. McMillan. Image-based visual hulls. *ACM Transactions on Graphics*, (6):369–374, 2000.
17. Meshlab. <http://meshlab.sourceforge.net/>.
18. Y. Oki, M. Visentini-Scarzanella, T. Wada, R. Furukawa, and H. Kawasaki. Entire shape scan system with multiple pro-cams using texture information and accurate silhouette creating technique. In *Proceedings of The 14th IAPR International Conference Machine Vision Applications*, pages 18–21, 2015.
19. S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2006*, pages 519–528, 2006.
20. The Computational Geometry Algorithms Library. <http://www.cgal.org/>.
21. Y. Thibault, H. Kawasaki, R. Sagawa, and R. Furukawa. Exemplar based texture recovery technique for active one shot scan. In *Proceedings of The 13th IAPR Conference on Machine Vision Applications*, pages 331–334, 2013.
22. K.-J. Yoon and I. S. Kweon. Adaptive support-weight approach for correspondence search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(4):650–656, 2006.