PAPER

3D Reconstruction with Globally-Optimized Point Selection

Norimichi UKITA[†a)], ***Senior Member and*** **Kazuki MATSUDA**[†], ***Nonmember***

SUMMARY This paper proposes a method for reconstructing accurate 3D surface points. To this end, robust and dense reconstruction with Shape-from-Silhouettes (SfS) and accurate multiview stereo are integrated. Unlike gradual shape shrinking and/or bruteforce large space search by existing space carving approaches, our method obtains 3D points by SfS and stereo independently, and then selects correct ones from them. The point selection is achieved in accordance with spatial consistency and smoothness of 3D point coordinates and normals. The globally optimized points are selected by graph-cuts. Experimental results with several subjects containing complex shapes demonstrate that our method outperforms existing approaches and our previous method.

key words: *shape reconstruction, shape from silhouettes, multiview stereo, graph-cuts*

1. Introduction

3D reconstruction from multiple views is an important issue in computer vision. Our method employs two types of 3D reconstruction techniques, namely shape-from-silhouettes (SfS) and multiview stereo, which are widely used for camera-based 3D reconstruction.

In SfS, multiview silhouettes of a target object are projected to a 3D space, and their intersection is regarded as the volume of the object, which is called a visual hull. While SfS is fast, robust, and able to get dense and smooth points, the visual hull might include false-positives in the concave regions of the object shape as shown in Figs. 1 and 2.

In multiview stereo, image windows that match between multiple views are found with photo-consistency in order to compute the distance to the 3D point of interest. In principle, every point where multiview matching is established can be reconstructed. Difficulty in matching is caused in shaded, textureless, and uniquely textured regions. This difficulty results in sparse and incorrect 3D points.

This paper proposes how to integrate the advantages of the above two schemes. In the proposed approach, 3D points are selected from the results of the two schemes so that a visual hull from SfS is partly replaced by a point cloud reconstructed by multiview stereo. This point selection replaces false-positive points in the visual hull by true-positive points in the point cloud of multiview stereo.

Since a lot of existing algorithms gradually refine a *small* range of the surface of the visual hull in an iterative manner, they tend to have local optima in iteration (e.g. see [1]). While recent advances in optimization techniques allow us to acquire a globally optimal shape from a whole *large* space where the real shape of an observed object possibly exists, global optimization in the large search space requires huge computational cost (e.g. around an hour or more in [3], [11]). Our approach resolves the problems of these previous methods by point selection only from the visual hull surface and the stereo point cloud, not from all possible points in the large space. This is because:

- The stereo point cloud is globally optimized because multiview stereo reconstructs it from all possible combinations of feature points extracted from images. While the point cloud might be incomplete due to lack of the feature points on the surface of the observed object, reconstruction of a limited number of the feature points can be finished efficiently.

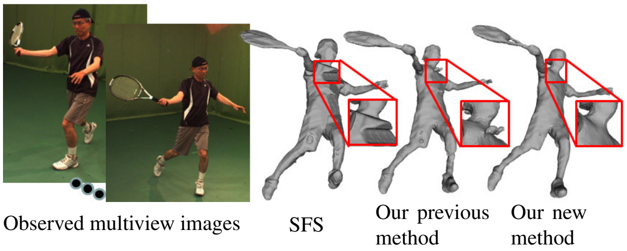


Fig. 1 Our new result compared with the results of Shape from Silhouette and our previous method [21]. For visualization purpose, a 3D surface mesh was generated from reconstructed 3D points. Typical differences are seen in the regions enclosed by rectangles. SfS produced huge error in concave regions, which were occluded by the right arm. A protrusion, which was generated due to a small number of extraneous 3D points, was remained in the same region by our previous method. To remove those 3D points by global optimization is the goal of our new method.

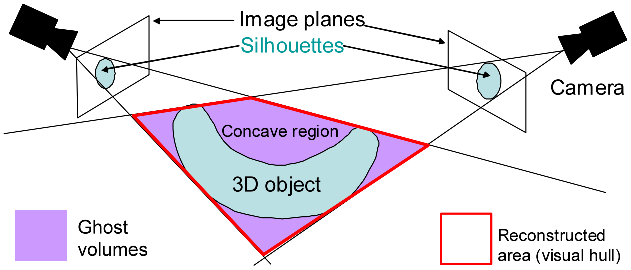


Fig. 2 Shape from silhouettes.

Manuscript received February 27, 2012.
Manuscript revised July 9, 2012.

[†]The authors are with the Graduate School of Information Science, Nara Institute of Science and Technology, Ikoma-shi, 630–0192 Japan.
a) E-mail: ukita@is.naist.jp

DOI: 10.1587/transinf.E95.D.3069

- Incomplete regions of the stereo point cloud are filled by a part of the visual hull, which can be reconstructed much faster than other 3D reconstruction techniques. It should be noted that even global optimization of the whole space such as [2], [3], [11] cannot reconstruct the featureless regions.

In practice, our point selection between 3D points reconstructed by SfS and multiview stereo is achieved using global optimization by graph-cuts with smoothness constraints and penalty distance between visual hull points and stereo points. As shown in Fig. 1, this globally optimized point selection outperforms our previous method based on local point selection with naive thresholding [21].

2. Related Work

2.1 Shape-from-Silhouettes

Figure 2 illustrates a visual hull reconstructed by SfS [4]. Even if the multiview silhouettes of a target object are extracted correctly, the visual hull might include false-positives as well as the real shape of the object. The false-positives are called *ghost volumes*. While they are reduced as the number of the cameras grows, it is impossible to remove them in the concave regions of the object.

Despite the ghost volumes, SfS is widely used for shape reconstruction in a studio. This is because SfS can obtain dense and smooth surface points, and silhouette extraction is easier than stereo point correspondence in an experimental environment such as a chroma key studio.

Silhouette constraints have the advantage also that they provide 3D points on occluded surfaces along the silhouettes (e.g. boundaries between the torso and the sleeve in the lefthand image of Fig. 4).

2.2 Multiview Stereo

Although early works in multiview stereo match all points independently, recent approaches find the points on the surface that minimizes a global photo-consistency with smoothness constraints (e.g. optimized by level sets [12], [13], and EM [14]). Novel techniques can reconstruct normals as well as 3D points; for example, [15], [16].

While multiview stereo can reconstruct accurate 3D positions, it cannot reconstruct textureless regions, which make point correspondence difficult. This difficulty causes incorrect and/or incomplete surface reconstruction. 3D points on occluded surfaces also cannot be reconstructed.

Photo-consistency becomes more powerful with artificial textures (e.g. chessboard patterns with multiple colors) on a target surface [17], [18]. However, such specially-colored textures are unavailable in natural scenarios.

To see the typical difference between SfS and multiview stereo, the results given by these methods are shown in Fig. 4, which were obtained from images shown in Fig. 3. For visualization purpose, the mesh surfaces obtained from the reconstructed 3D points, each of which was obtained by Poisson surface reconstruction [19], are shown in Fig. 4. For emphasizing the limitations of each method, specially-colored clothing was used for robust silhouette extraction and multiview point matching. While the visual hull produces the feasible surface with no missing body-regions, several body-regions (e.g. feet) are missing in the one reconstructed from the stereo point cloud. This is because no point correspondence was obtained in these regions. From the stereo point cloud, on the other hand, thin sleeves are reconstructed correctly although ghost volumes make the sleeves thicker than the real shape.

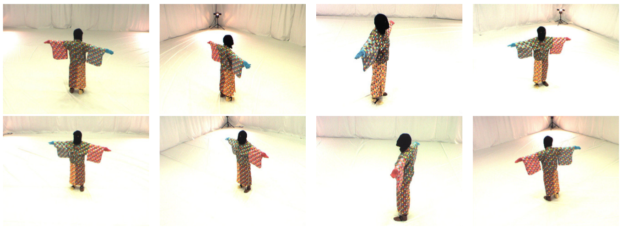
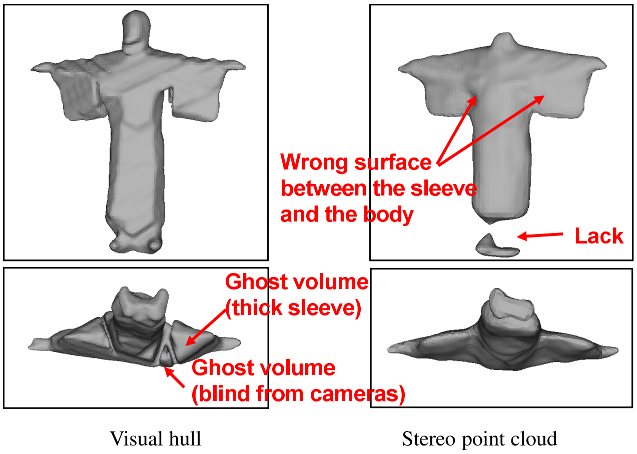


Fig. 3 Images captured from eight viewpoints and used for reconstructing 3D surfaces shown in Fig. 4.



Visual hull

Stereo point cloud

Fig. 4 Mesh surfaces obtained from a visual hull and a stereo point cloud, which were reconstructed from multiview images shown in Fig. 3.

2.3 Space Carving and Its Variations

The most popular approach for refining a visual hull is space carving [1]. The visual hull, which is an initial shape, is carved until photo-consistency is satisfied between multiple views. Other constraints such as smoothness can be also optimized (e.g. using continuous local optimization by gradient descent [6], discrete global optimization by graph cuts [7], [8], [10], and continuous global optimization [9]). Furthermore, bruteforce optimization of a large space within the visual hull [2], [3], [11] can avoid local optima, which appear between the visual hull and the real surface.

A series of space carving have the limitations below:

- Gradual carving from a visual hull tends to fall into local optima.

- Optimization of the whole large space requires huge computational cost; around an hour or more.

3. Optimized Selection of 3D Points

Existing approaches in space carving refines surface points in a visual hull until photo-consistency is satisfied. Photo-consistency could be satisfied before the reconstructed surface reaches the real surface.

Instead of carving the whole large space inside the visual hull [3], [11], our previous method [21] reconstructs surface points by multiview stereo [15] and SfS independently, and then combines the segments of the surface points so that the surface of the visual hull that occludes the stereo points are carved. The numbers of 3D points evaluated in the carving process are $O(r^3)$ and $O(r^2)$, where r denotes the radius of a target object, in carving the whole visual hull [3], [11] and carving the surface of the visual hull [21], respectively.

While our previous method carves ghost volumes efficiently, sensitive thresholding in carving might miss-carve the ghost volumes and/or over-carve the visual hull. A method proposed in this paper resolves these problems by globally optimized point selection from the surface points reconstructed by multiview stereo and SfS.

After introducing a basic algorithm of visual hull carving in our previous method in Sect. 3.1, Sect. 3.2 and 3.3 describe optimized carving of the visual hull using graph-cuts and pruning unreliable stereo points, respectively. Our optimized point selection is achieved with those visual hull carving and stereo point pruning.

3.1 Local Point Carving with Naive Thresholding

First of all, SfS and multiview stereo are applied to images independently. Only the surface points of the reconstructed visual hull are remained as illustrated in Fig. 5.

Then, like ray tracing, each of stereo points (e.g. s_1 and s_2 in Fig. 6) is projected onto image planes in which the point is extracted along the ray (e.g. $r_{1,1}$ *and* $r_{1,2}$ in Fig. 6, where $r_{s,c}$ denotes a ray from s-th stereo point to c-th camera center). The rays are drawn from every stereo point to camera centers, each of which observes that stereo point; a stereo point is observed from a camera, if the point is reconstructed using the image of the camera in multiview stereo. If the ray hits one or more points in the visual hull (e.g. v_1, v_3, and v_4 in Fig. 6), these points are carved as ghost volumes. In practice, the bounding box around each point of the visual hull (e.g. bounding boxes b_1 and b_4 around v_1 and v_4 in Fig. 6) is prepared for this intersection test [20]. If the ray crosses the box, its respective visual hull point is carved.

For this intersection test, the size of the bounding box is critical. If the size is smaller/larger, visual hull points that must be carved/remained are remained/carved incorrectly. The size is determined in accordance with the distance between the stereo point of interest and its nearest stereo point.

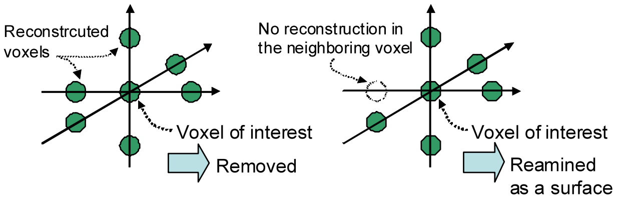


Fig. 5 Surface points extraction from voxels in a visual hull.

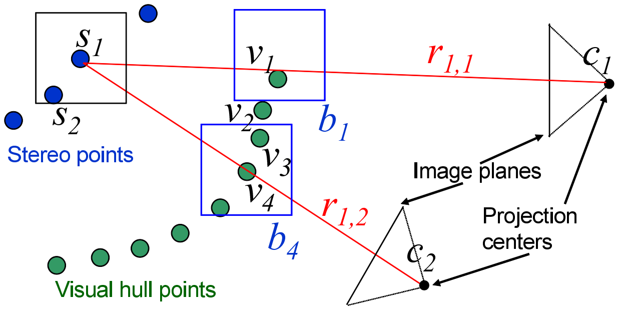


Fig. 6 Shape carving by testing ray intersection with a bounding box.

Figure 6 illustrates this intersection test. Given a stereo point, s_1 and its nearest point, s_2, the size of the bounding box for carving along rays drawn from s_1 is equal to that of the cube whose center is s_1 and that passes s_2. This bounding box is located in every visual hull point. Assume that s_1 is reconstructed by cameras 1 and 2, whose projection centers are c_1 and c_2, respectively. Since the ray from s_1 to c_1 (denoted by $r_{1,1}$) passes through b_1, v_1 is carved. v_2 is also carved. Similarly, v_3 and v_4 are also carved because their bounding boxes are on the way of $r_{1,2}$.

3.2 Optimized Carving of Visual Hull by Stereo Points

As with our previous method [21] mentioned in Sect. 3.1, our new method also performs SfS and multiview stereo independently.

Point carving in our previous method might generate non-smooth surfaces with holes and extraneous points because of the following reasons:

Naive size selection of bounding-boxes: A visual hull point is remained even a little outside a bounding box, or removed even a little inside a bounding box.

Independent local carving: Each visual hull point is carved independently of whether or not its neighboring visual hull points are carved.

To resolve these problems, our new method carves the surface of the visual hull so that the surface is globally optimized in terms of “proximity between a visual hull point and carving rays” and “smoothness of surface points”. This optimization is achieved by the following penalty functions (see Fig. 7):

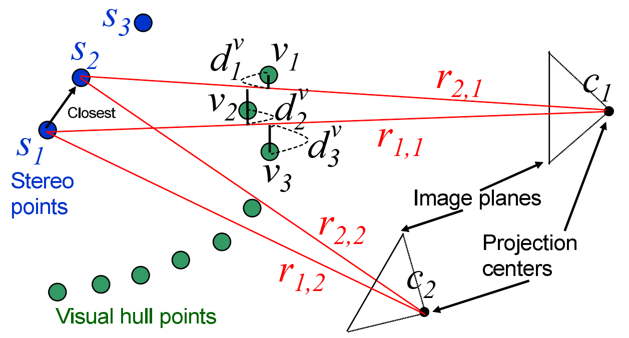


Fig. 7 Carving surface points of a visual hull. The surface points are carved if they occlude points obtained by multiview stereo. Occlusion check is achieved based on the distance between a surface point of interest and a ray from a stereo point to a camera; if the surface point is close to the ray, the point is carved.

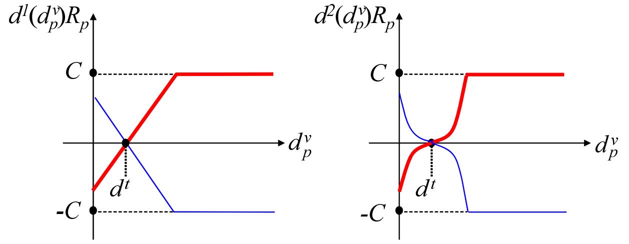


Fig. 8 Penalty functions. While $d^1(p)$ is a linear function, $d^2(p)$ gives a much larger penalty as d_p^v gets away from d^t. Thin blue and thick red lines show penalty values if v_p is remained (i.e $R_p = -1$) and removed (i.e. $R_p = 1$), respectively.

$$P_D = \sum_{p}^{N^v} d(d_p^v) R_p, \tag{1}$$

$$P_S = \sum_{p}^{N^v} \sum_{n \in \boldsymbol{V}_p} ||R_p - R_n||, \tag{2}$$

where

- N^v denotes the number of the surface points extracted from the visual hull,
- $d(d_p^v)$ is a distance function that evaluates the need to remain p-th surface point, v_p, where d_p^v denotes a distance from v_p to its closest ray (e.g. d_1^v is the distance between v_1 to $r_{2,1}$, d_2^v is between v_2 to $r_{1,1}$, and d_3^v is between v_3 to $r_{1,1}$ in Fig. 7),
- $\boldsymbol{V}_p$ includes at most six neighboring surface points (i.e. upper, lower, left, right, front, and back points)[†] of v_p, and
- $R_p \in \boldsymbol{R} = \{R_1, \cdots, R_{N^v}\}$ has -1 or 1. If R_p has $-1/1$, p-th surface point is remained/removed. Initially, all R_p is set to -1.

In our experiments, the following two kinds of distance functions are tested, namely a L1-norm-based distance function, $d^1(p)$, and a L2-norm-based distance function, $d^2(p)$ (see Fig. 8):

$$d^1(d_p^v) = \min(d_p^v - d^t, C), \tag{3}$$

$$d^2(d_p^v) = \text{sgn}(d_p^v - d^t) \min((d_p^v - d^t)^2, C), \tag{4}$$

where:

- If and only if $d_p^v = d_r^t$, a penalty value given to v_p is 0 whether v_p is removed or remained. d_r^t, which is used for rays drawn from r-th stereo point s_r, is determined so that these rays and the ones drawn from stereo points close to s_r carve ghost volumes without missing. Specifically, given stereo point s_r that draws the ray closest to v_p, d_r^t is equal to the length between s_r and its closest stereo point.
- C $(= d(3d^t)$ in our experiments) is a constant for cutoff.
- $\text{sgn}(x)$ is a sign function.

While the penalty function (1) evaluates the proximity only with the closest ray, it can be evaluated with multiple rays for robust evaluation. The penalty function (1) is rewritten as follows:

$$P_{DM} = \sum_{p}^{N^v} \left(\sum_{q \in \boldsymbol{Q}_p} w(d(d_p^v, q)) d(d_p^v, q) \right) R_p, \tag{5}$$

where

- $\boldsymbol{Q}_p$ is a set of carving rays that are the top N closest ones to v_p,
- $d(d_p^v, q)$ denotes the distance between v_p and q-th ray in $\boldsymbol{Q}_p$, and
- $w(d(d_p^v, q))$ is a weighting variable for $d(d_p^v, q)$, where $w(d(d_p^v, q)) = \exp(-d(d_p^v, q))$.

In the formulation described above, R_p is optimized so that the weighted sum of P_{DM} and P_S is minimized:

$$w_D P_{DM} + w_S P_S, \tag{6}$$

where w_D and w_S are weighting variables. These variables are determined so that $w_S / w_D = C$. The weighted sum (6) is globally minimized by using graph-cuts [22].

3.3 Pruning Stereo Points

Our method employs PMVS [15] as multiview stereo, which is top-ranked in the Middlebury database [5]. While PMVS could get better results, it still has the problems below:

- Pixels along an object boundary in images are aggressively used for 3D reconstruction. While this process increases the number of the 3D points, correct matching with these pixels is difficult because of background pixels contained in an image window used for matching. Matching error produces inaccurate 3D points.
- Accuracy of the normal gets lower where multiview stereo reconstructs the point with pixels along an object boundary in images.

[†] Since all surface points are extracted from 3D grid voxels, it is known that whether any pair of surface points are neighbors in a 3D grid space.

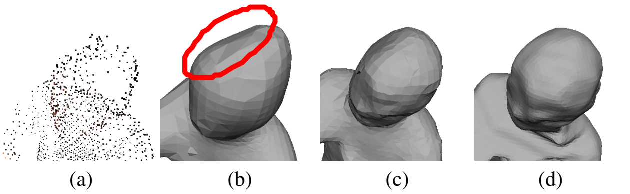


Fig. 9 Reconstruction error of multiview stereo [15]. (a) 3D points reconstructed by multiview stereo. (b) 3D surface reconstructed from (a). A part of the head was dented. (c) 3D surface reconstructed using (a) by our method. The dent is still remaining. (d) 3D surface reconstructed by our method with stereo point pruning described in Sect. 3.3.

Compared with these problematic points, the surface of a visual hull is reconstructed robustly. Figure 9 (a), (b), and (c) show examples.

To resolve the problems mentioned above, stereo points with the following properties are pruned from the result of multiview stereo if they are near the surface of a visual hull: 1) the point is reconstructed by pixels around the boundary of a silhouette and 2) the normal is significantly different from the normal of the nearest visual hull surface. Specifically, if the distance from a stereo point to its nearest visual hull surface is shorter than a threshold, t_s, and either of the following conditions is satisfied, the stereo points is pruned:

1. The stereo point is projected onto all image planes used for reconstructing the point. Then the distance between the projected pixel and the boundary of a silhouette is less than a threshold, t_b, at least in one of the images.
2. The angle between the normals of the stereo point of interest and the nearest visual surface is larger than a threshold, t_θ.

In our experiments, t_s is the side length of a voxel, t_b is the side length of an image window used for matching in multiview stereo, and $\theta_a = 30$ degrees. It should be noted that two thresholds t_s and t_b are determined automatically in accordance with the spatial resolution of reconstruction.

The remaining stereo points are used for optimized carving of a visual hull, as described in Sect. 3.2. Figure 9 (d) shows the reconstructed surface from the 3D points acquired by optimized carving.

4. Experiments

The proposed method was applied to multiview image sequences for validating the effectiveness of the method. All results were obtained from eight cameras. The cameras were located around a subject. If a camera was located right above the subject, it could reduce ghost volumes by SfS, especially those surrounded by arms. But no camera was located above the subject for verifying the performance of carving the ghost volumes under severe conditions.

3D surfaces reconstructed using L1-based-norm and L2-based-norm distance functions are shown in Fig. 10. While two results were almost same, the L2-based-norm distance function overcarved the leg (enclosed by a circle in

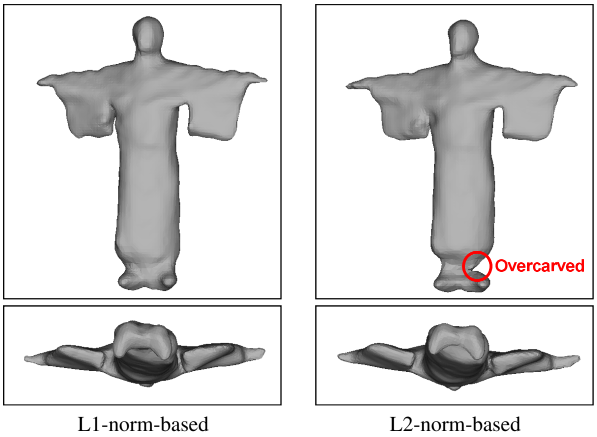


Fig. 10 Reconstructed surfaces using different distance functions, namely the L1-norm-based function in expression (3) and the L2-norm-based function in expression (4).

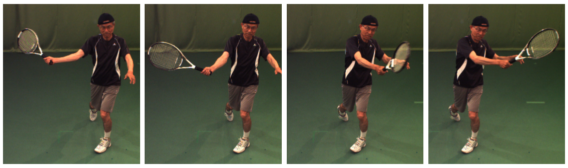


Fig. 11 Tennis sequence: observed images.

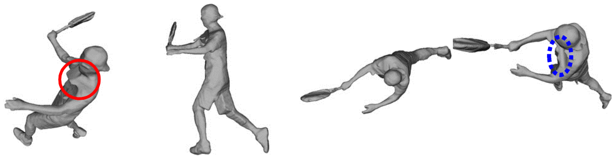


Fig. 12 Tennis sequence: shape from silhouettes.

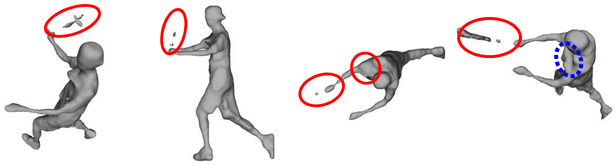


Fig. 13 Tennis sequence: space carving [2].

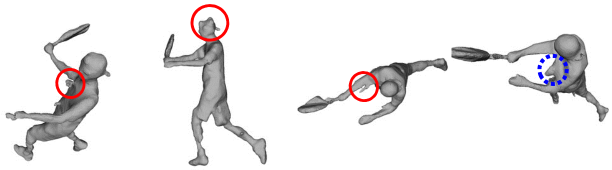


Fig. 14 Tennis sequence: our previous method [21].

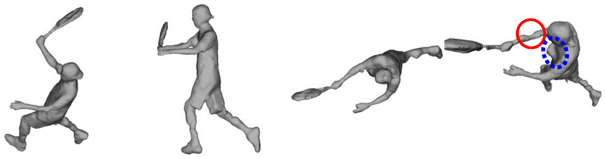


Fig. 15 Tennis sequence: our method.

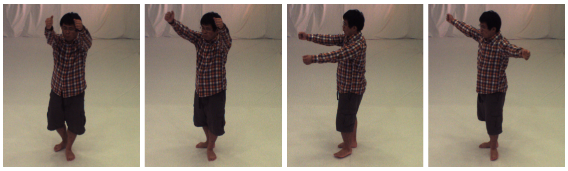


Fig. 16 Exercise sequence: observed images.

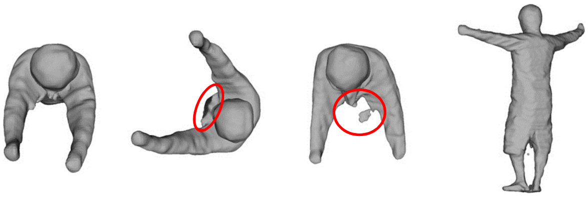


Fig. 17 Exercise sequence: shape from silhouettes.

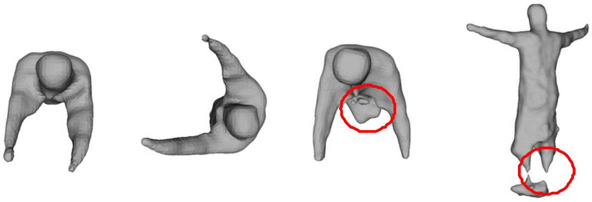


Fig. 18 Exercise sequence: space carving [2].

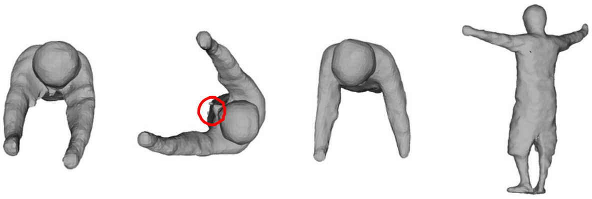


Fig. 19 Exercise sequence: our previous method [21].

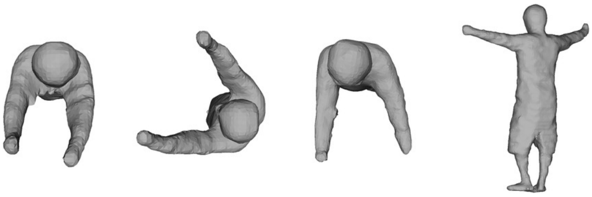


Fig. 20 Exercise sequence: our method.

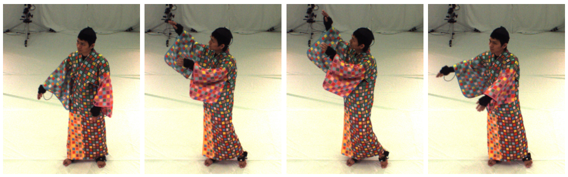


Fig. 21 Dance sequence: observed images.

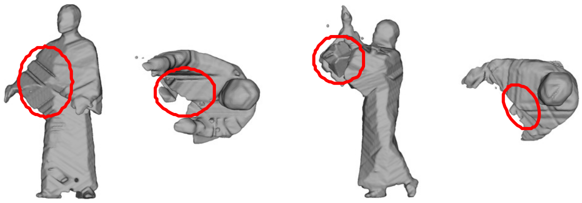


Fig. 22 Dance sequence: shape from silhouettes.

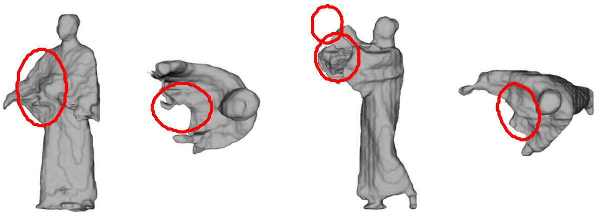


Fig. 23 Dance sequence: space carving [2].

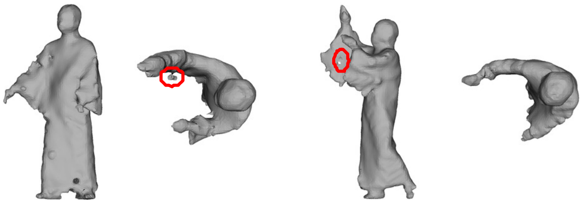


Fig. 24 Dance sequence: our previous method [21].

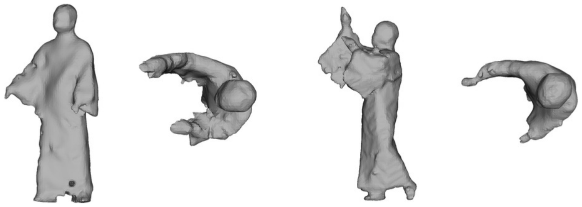


Fig. 25 Dance sequence: our method.

the figure). In accordance with these results, the L1-based-norm distance was employed in our experiments, though it seems that the distance function should be selected depending on images and/or objects.

Figures 11–35 show observed images and the results of surface reconstruction from them; playing tennis (Figs. 11–15), exercising (Figs. 16–20), dancing with loose-fitting clothing (Figs. 21–25), throwing (Figs. 26–30), and batting sequences (Figs. 31–35). From each sequence, four sets of images and reconstructed shapes observed at different moments are shown.

For comparison, the results of shape-from-silhouettes (Figs. 12, 17, 22, 27, and 32), space carving with graph-cuts [2] (Figs. 13, 18, 23, 28, and 33), our previous method [21] (i.e. shape carving without optimized point selection by graph-cuts) (Figs. 14, 19, 24, 29, and 34), and our new method (Figs. 15, 20, 25, 30, and 35) are shown. All parameters in SfS, multiview stereo, and surface reconstruction were same in all methods. The numbers of surface voxels and stereo points were around 50000–70000 and 3000–6000, respectively, at each frame. Note that the number of nodes in carving with graph-cuts was equal to that of

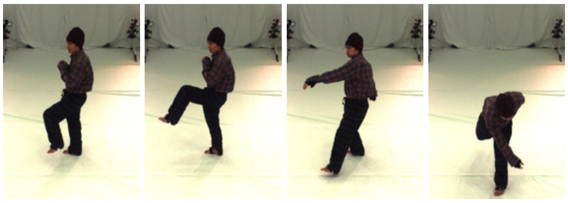


Fig. 26 Throwing sequence: observed images.

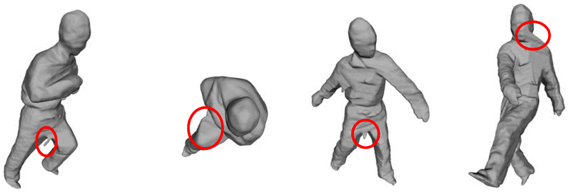


Fig. 27 Throwing sequence: shape from silhouettes.

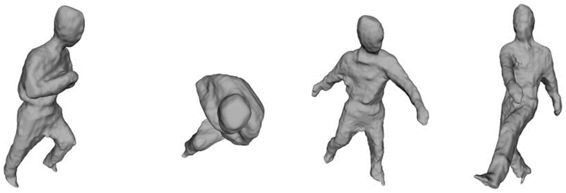


Fig. 28 Throwing sequence: space carving [2].

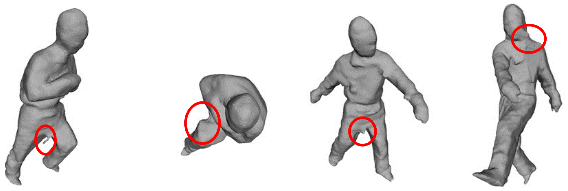


Fig. 29 Throwing sequence: our previous method [21].

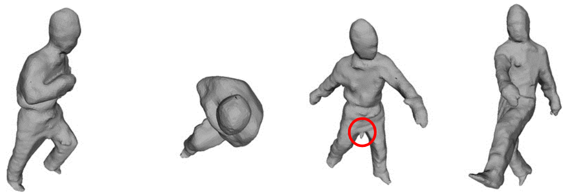


Fig. 30 Throwing sequence: our method.

the surface voxels (i.e. 50000–70000).

In our method, the optimized 3D points were acquired around one minute by Xeon 2.4 GHz: 10 seconds in multiview stereo, a few seconds in pruning stereo points, and 30–60 seconds in carving using graph-cuts. This computational cost is significantly smaller than that of global optimization in the whole large space search [2], [3], [11], where the number of nodes in graph-cuts is upto 20 million [11]; our method was around 200 times faster than [3].

Solid and dotted circles in the figures indicate "error regions that were correctly carved by one or more of other methods" and "error regions where all methods could not get correct shapes", respectively. Errors enclosed by the dotted circles were observed in severe concave regions surrounded by the arms. Other errors and properties of the methods are summarized as follows:

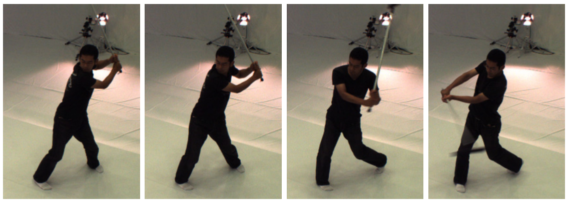


Fig. 31 Batting sequence: observed images.

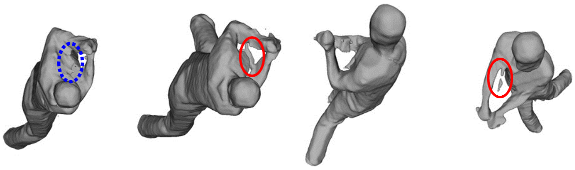


Fig. 32 Batting sequence: shape from silhouettes.

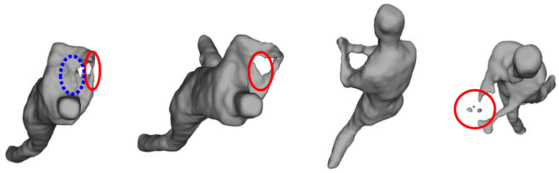


Fig. 33 Batting sequence: space carving [2].

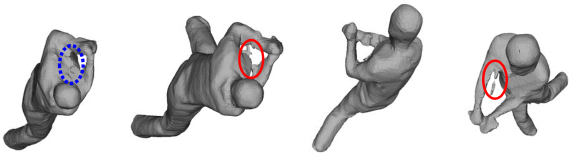


Fig. 34 Batting sequence: our previous method [21].

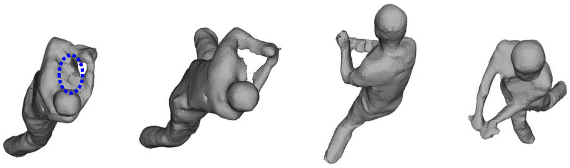
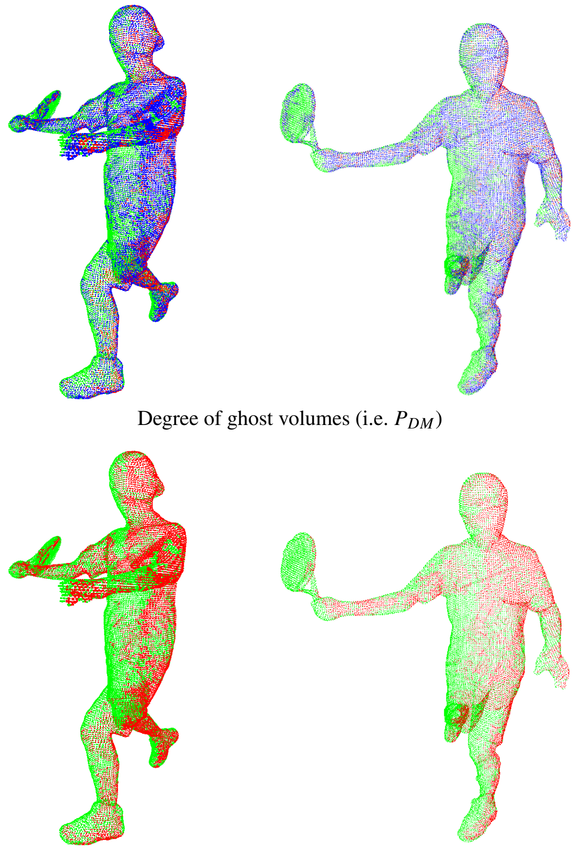


Fig. 35 Batting sequence: our method.

- Shape-from-silhouettes produced large ghost volumes in all of concave regions.
- Space carving sometimes stopped before reaching a real surface (e.g. regions between the arms in Fig. 23 and the third image from the left 18) and overcarved a slim region (e.g. a tennis racket in Fig. 13 and the arm in the rightmost image in Fig. 33).
- While shape carving reduced ghost volumes rather than space carving, some small protrusions were re-



Degree of ghost volumes (i.e. P_{DM})

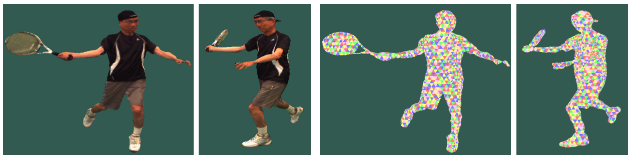
Carved and remaining points

Fig. 36 Tennis sequence: A visual hull observed from different two viewpoints is shown. In upper images (i.e. the degree of ghost volumes), P_{DM} of each point was higher in order of green, blue, and red. In bottom images, red voxels were carved while green voxels remained.

mained. These protrusions were caused due to surface points of a visual hull that could not be removed. In addition, shape deformation was caused due to 3D oriented points incorrectly reconstructed by PMVS; for example, the head was overcarved in the second image from the left in Fig. 14, as with the one in Fig. 9 (b) and (c).

- In all of the results, our method could get plausible shapes with less noticeable errors than other methods, except a small protrusion in the third result from the left in Fig. 30, which was not seen in the shape reconstructed by space carving [2] as shown in Fig. 28.

For further intuitive understanding of optimization in the proposed method, the example of an optimization result is shown in Fig. 36, where penalty values P_{DM} in a visual hull (in upper images) and carved and remaining visual hull points in the optimization result (in bottom images). It can be seen that the carved points were connected to each other so that they made clusters, while several isolate points had larger penalty values, P_{DM}. This effect was acquired by smoothness term (2).

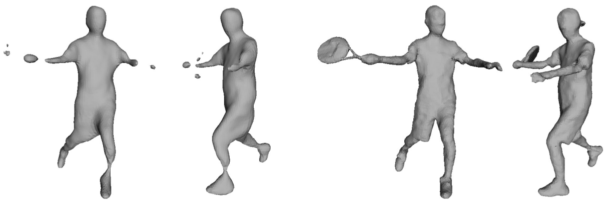


(a) Surface textured by real images (b) Surface textured by color patches

Fig. 37 Textured 3D surfaces, which are identical in shape, used for quantitative comparison. Two surfaces observed from different two viewpoints are shown.

Table 1 RMS error comparison of five methods, SfS, PMVS [15], space carving with graph cuts [2], our previous method [21], and the proposed method.

	SfS	PMVS [15]	SC [2]	Previous [21]	Proposed
(1)	11.2	13.3	9.7	7.3	6.9
(2)	11.2	3.6	8.0	4.4	4.1



(a) Surface reconstructed from images capturing the surface textured by real images (Fig. 37 (a))

(b) Surface reconstructed from images capturing the surface textured by color patches (Fig. 37 (b))

Fig. 38 3D surfaces reconstructed by PMVS [15]. Two surfaces observed from different two viewpoints are shown.

Reconstruction accuracy was evaluated with two synthesized 3D surfaces (Fig. 37). The two surfaces had the same shape generated as follows: i) the initial shape was generated by the proposed method with eight real cameras and ii) the initial shape was rectified manually so that it got close to the real shape. Then the surface was textured by simple OpenGL functions with two kinds of patterns, namely (1) captured images used for reconstructing the surface and (2) random color patches, which are useful for window matching in multiview stereo. These two textured surfaces were separately observed from eight viewpoints (in a simulation environment), whose geometric configuration relative to the target surface was different from that of the real eight cameras, in order to capture the multiview images of each surface. The captured images were then used for 3D reconstruction. The reconstruction accuracy was evaluated by a distance from each surface point to its nearest reconstructed point. Table 1 shows the RMS errors of the distance. The RMS errors were compared among five methods. From the table, the following observations can be seen:

(1) For reconstruction of the less-textured surface shown in Fig. 37 (a), the proposed method outperformed the others. It should be noted that PMVS [15] got the worst result because several body regions were not reconstructed due to poor textures, as shown in Fig. 38 (a).

(2) With rich textures shown in Fig. 37 (b), PMVS got the best result as shown in Fig. 38 (b), followed by the proposed method. This is because several stereo points were over-pruned in the proposed method.

As the summary of the above observations, the proposed method is considered to be superior in reconstructing less-textured surfaces, while multiview stereo can obtain more accurate results if a target surface has rich textures. Note that, even if the target surface has rich textures, the proposed method is comparable to multiview stereo.

5. Concluding Remarks

This paper proposed 3D point reconstruction from multiple views. The method employs two kinds of point sets reconstructed by SfS and multiview stereo. For sorting out these two kinds of the point sets, a two-phased point removal technique is proposed: 1) pruning of stereo points based on "proximity between their respective pixels and the object boundary in images" and "irregularity of the point normals" and 2) globally optimized carving of the surface of a visual hull by using the stereo points with graph-cuts.

The codes of PMVS [15], Poisson Surface reconstruction [19], and space carving with graph-cuts [2] were given by Y. Furukawa, M. Kazhdan, and S. Nobuhara, respectively.

References

[1] K.N. Kutulakos and S.M. Seitz, "A theory of shape by space carving," IJCV, vol.38, no.3, 2000.
[2] T. Tung, S. Nobuhara, and T. Matsuyama, "Simultaneous super-resolution and 3D video using graph-cuts," CVPR, 2008.
[3] C. Hernandez, G. Vogiatzis, and R. Cipolla, "Probabilistic visibility for multi-view stereo," CVPR, 2007.
[4] G. Cheung, T. Kanade, J. Bouguet, and M. Holler, "A real time system for robust 3D voxel reconstruction of human motions," CVPR, 2000.
[5] S.M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A comparison and evaluation of multi-view stereo reconstruction algorithms," CVPR, 2006.
[6] C.H. Esteban and F. Schmitt, "Silhouette and stereo fusion for 3D object modeling," CVIU, vol.96, no.3, 2004.
[7] S. Tran and L. Davis, "3d surface reconstruction using graph cuts with surface constraints," ECCV, 2006.
[8] G. Vogiatzis, P.H. Torr, and R. Cipolla, "Multi-view stereo via volumetric graph-cuts," CVPR, 2005.
[9] K. Kolev, M. Klodt, T. Brox, and D. Cremers, "Continuous global optimization in multiview 3D reconstruction," IJCV, vol.84, no.1, pp.80–96, 2009.
[10] S. Sinha and M. Pollefeys, "Multi-view reconstruction using photo-consistency and exact silhouette constraints: A maximum-flow formulation," ICCV, 2005.
[11] V. Lempitsky, Y. Boykov, and D. Ivanov, "Oriented visibility for multiview reconstruction," ECCV, 2006.
[12] O. Faugeras and R. Keriven, "Variational principles, surface evolution, PDE's, level set methods and the stereo problem," IEEE Trans. Image Process., vol.7, no.3, pp.336–344, 1998.
[13] J.-P. Pons, R. Keriven, and O.D. Faugeras, "Modelling dynamic scenes by registering multi-view image sequences," CVPR, 2005.
[14] C. Strecha, R. Fransens, and L.V. Gool, "Combined depth and outlier estimation in multi-view stereo," CVPR, 2006.
[15] Y. Furukawa and J. Ponce, "Accurate, dense, and robust multiview stereopsis," IEEE Trans. Pattern Anal. Mach. Intell., vol.32, no.8, pp.1362–1376, 2010.
[16] N. Ahmed, C. Theobalt, P. Dobrev, H.-P. Seidel, and S. Thrun, "Robust fusion of dynamic shape and normal capture for high-quality reconstruction of time-varying geometry," CVPR, 2008.
[17] V. Scholz, T. Stich, M. Keckeisen, M. Wacker, and M. Magnor, "Garment motion capture using color-coded patterns," Comput. Graph. Forum, vol.24, no.3, pp.439–448, 2005.
[18] R. White, K. Crane, and D. Forsyth, "Capturing and animating occluded cloth," ACM TOG (SIGGRAPH), vol.26, no.3, 2007.
[19] M. Kazhdan, M. Bolitho, and H. Hoppe, "Poisson surface reconstruction," SGP, 2005.
[20] T. Moller and B. Trumbore, "Fast, minimum storage ray-triangle intersection," J. Graphic Tools, vol.2, no.1, pp.21–28, 1997.
[21] K. Matsuda and N. Ukita, "Direct shape carving: Smooth 3D points and normals for surface reconstruction," IEICE Trans. Inf. & Syst., vol.E95-D, no.7, pp.1811–1818, July 2012.
[22] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," IEEE Trans. Pattern Anal. Mach. Intell., vol.23, no.11, pp.1222–1239, 2001.

Norimichi Ukita received the Ph.D degree in Informatics from Kyoto University, Japan, in 2001. After working as an assistant professor at Nara Institute of Science and Technology (NAIST), he became an associate professor in 2007. He was a research scientist of PRESTO, Japan Science and Technology Agency (JST) from 2002 to 2006, and a visiting research scientist at the Robotics Institute, Carnegie Mellon University from 2007 to 2009. He is now working also as a visiting researcher at the Intelligent Robotics and Communication Laboratories, ATR. His main research interests are object detection/tracking and human pose/shape estimation. He has received the best paper award from the IEICE in 1999.

Kazuki Matsuda was a master course student of Nara Institute of Science and Technology. His research theme was 3D shape reconstruction.