

# Shape Reconstruction with Globally-optimized Surface Point Selection

Norimichi Ukita, Kazuki Matsuda, and Norihiro Hagita  
Nara Institute of Science and Technology  
ukita@ieee.org

## Abstract

*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 brute-force 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 demonstrate that our method outperforms existing approaches.*

## 1. Introduction

3D reconstruction from multiviews is an important issue in computer vision. Our method integrates two types of 3D reconstruction techniques, namely shape-from-silhouettes (SfS) and multiview stereo, while utilizing their advantages.

Since a lot of similar 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; see [1], for example. While recent advance in optimization techniques allow us to acquire a globally optimal shape from a whole *large* space where the real shape of a target object possibly exists, global optimization in the large space requires huge computational cost (e.g. an hour or more in [5, 3]). Our approach resolves these problems by point selection only from the visual hull surface and the stereo point cloud.

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.

## 2. Related Work

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*.

Recent multiview stereo algorithms find the points on the surface that minimizes a global photo-consistency with smoothness constraints (e.g. optimized by level sets[6] and EM[7]). While multiview stereo can reconstruct accurate 3D positions, it cannot reconstruct textureless regions, which make point correspondence difficult.

The most popular approach for gradually refining a visual hull is space carving[1]. The visual hull, which is an initial shape, is carved until photo-consistency is satisfied between multiviews (e.g. [4]). Furthermore, brute-force optimization of a large space within the visual hull[5, 3, 2] can avoid local optima, which appear between the visual hull and the real surface. These kinds of space carving have the limitations below: gradual carving from a visual hull tends to fall into local optima, and optimization of the whole large space requires huge computational cost.

## 3. Optimized Selection of 3D Points

Unlike space carving, our previous method[10] reconstructs surface points by multiview stereo[8] and SfS independently, and then combines the segments of these 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[5, 3] and carving the surface of the visual hull[10], respectively. While our previous method[10] carves ghost volumes efficiently, sensitive thresholding in carving might miss-carve the ghost volumes and/or over-carve the visual hull.

### 3.1. Visual Hull Carving using Stereo Points

As with our previous method[10], our new method also performs SfS and multiview stereo independently.

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

**Naive thresholding:** A visual hull point is carved if it is proximal to a ray from a camera to a stereo point; see Fig. 1. This carving process is sensitive to a threshold for evaluating the proximity.

**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 it is globally optimized in terms of “proximity between a visual hull point and carving rays” and “smoothness of surface points”. These two criteria are evaluated by the following penalty functions (see Fig. 1):

$$P_D = \sum_p^{N^v} d(d_p^v) R_p, \quad (1)$$

$$P_S = \sum_p^{N^v} \sum_{n \in \mathbf{V}_p} \|R_p - R_n\|, \quad (2)$$

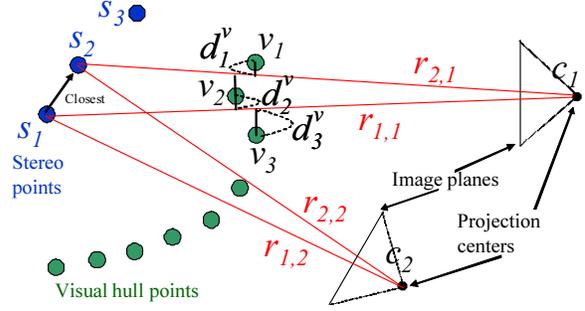
where

- $N^v$  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 keep  $p$ -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. 1),
- $\mathbf{V}_p$  includes at most six neighboring surface points (i.e. upper, lower, left, right, front, and back points) of  $v_p$ , and
- An unknown value  $R_p \in \mathbf{R} = \{R_1, \dots, R_{N^v}\}$  has -1 or 1 so that  $P_S$  is the Potts model[11]. If  $R_p$  is -1/1,  $p$ -th surface point is kept/removed. All values in  $\mathbf{R}$  are optimized in our method.  $R_p$  is initialized to be -1.

In our experiments, the following L1-norm based distance function is employed:

$$d(d_p^v) = \min(d_p^v - d^t, C), \quad (3)$$

where:



**Figure 1. Surface points of a visual hull 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  $d_p^v = d_r^t$ , a penalty value  $d(d_p^v)R_p$  given to  $v_p$  is 0 whether  $v_p$  is removed or kept.  $d(d_p^v)$  is negative/positive if  $d_p^v$  is less/more than  $d_r^t$ . Given stereo point  $s_r$  drawing the ray closest to  $v_p$ ,  $d_r^t$  is equal to the length between  $s_r$  and its closest stereo point.
- $C$  denotes a constant for cutoff.  $C$  was empirically determined to be  $d(3d^t)$ .

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:

$$P_{DM} = \sum_p^{N^v} \left( \sum_{q \in \mathbf{Q}_p} w(d(d_p^v, q)) d(d_p^v, q) \right) R_p \quad (4)$$

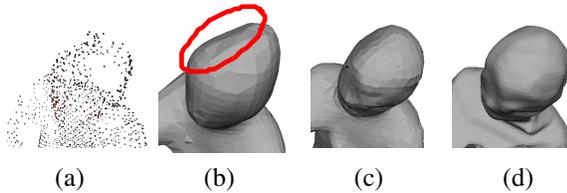
where

- $\mathbf{Q}_p$  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  $\mathbf{Q}_p$ , 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:

$$w_D P_{DM} + w_S P_S, \quad (5)$$

where  $w_D$  and  $w_S$  are weighting variables, which are determined so that  $w_S/w_D = C$ . The weighted sum (5) is globally minimized by using graph-cuts[12].



**Figure 2.** (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 still remains. (d) 3D surface reconstructed by our method with stereo point pruning.

### 3.2. Pruning Stereo Points

For multiview stereo, our method employs PMVS[8], which is fabulous but has the problems below. Since image patches along an object boundary are aggressively reconstructed, matching error with these pixels produces 3D points with inaccurate positions and normals. The normal must be accurate for surface patch reconstruction[9], which is used in our method. Bad effects of these problematic points are shown in Fig. 2 (a), (b), and (c).

To solve the above problems, our method prunes a stereo point if the distance from the stereo point to its nearest visual hull surface is shorter than a threshold,  $t_s$ , and either of the following conditions is satisfied:

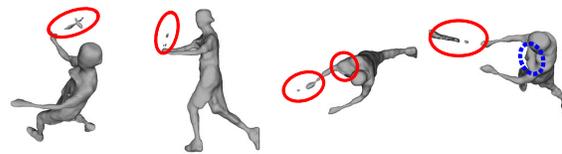
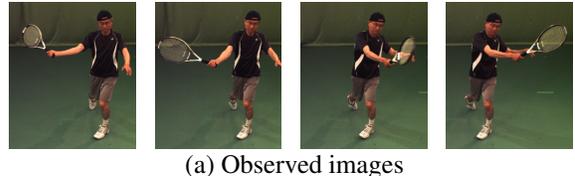
1. The stereo point is projected onto all image planes used for reconstructing the point. Then the distance between the projected pixel and the boundary of a silhouette is less than a threshold,  $t_b$  in at least one of the images.
2. The angle between the normals of the stereo point of interest and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

In our experiments,  $t_s$  is the side length of a voxel,  $t_b$  is the side length of an image patch used for matching in multiview stereo, and  $\theta_a = 30$  degrees.

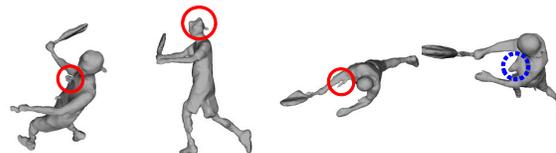
The remaining stereo points are used for optimized carving of a visual hull. Figure 2 (d) shows the effect of stereo point pruning mentioned here.

## 4. Experiments

Three kinds of sequences, which were captured by eight cameras, were used; see Fig. 3, 4, and 5. In each figure, four images and reconstructed shapes observed at different moments are shown. For comparison, the



(b) Results of space carving[2]



(c) Results of shape carving[10]



(d) Results of our method

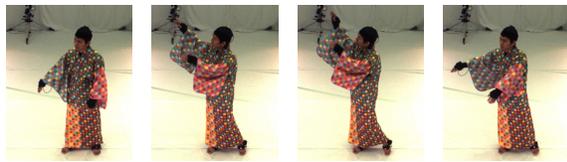
**Figure 3. Tennis sequence.**

results of space carving with graph-cuts[2], our previous method[10], and our new method are shown.

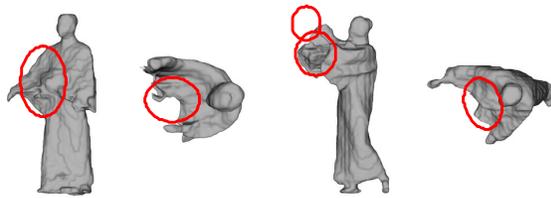
Our method obtained each shape in around one minute by Xeon 2.4GHz: 10 secs in stereo, a few secs in pruning stereo points, and 30–60 secs in carving with graph-cuts. This is around 200 times faster than [3].

Solid and dotted circles in the figures indicate “errors correctly carved by one or more of other methods” and “errors where all methods could not get correct shapes”, respectively. These errors are summarized as follows:

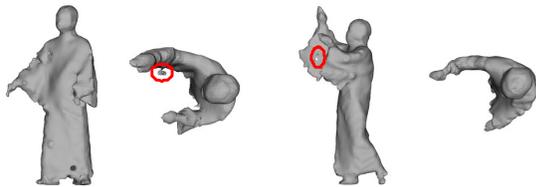
- Space carving sometimes stopped before reaching a real surface (e.g. regions between the the arms in Fig. 4 (b)) and overcarved a slim region (e.g. a racket in Fig. 3 (b) and the arm in Fig. 5 (b)).
- While most of local optima were avoided by shape carving, some small protrusions were left. In addition, 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. 3 (c).
- Our method could get plausible shapes with less noticeable errors than other methods.



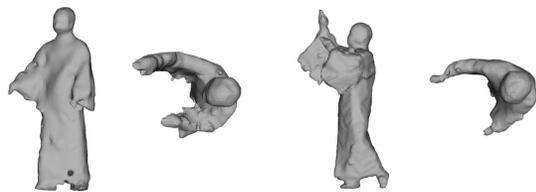
(a) Observed images



(b) Results of space carving[2]



(c) Results of shape carving[10]



(d) Results of our method

**Figure 4. Dance sequence where stereo can work well with dense textures.**

## 5. Concluding Remarks

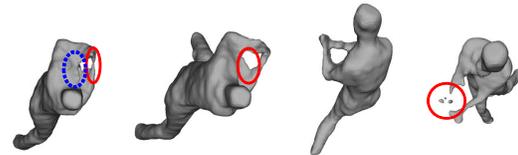
The proposed method efficiently integrates two kinds of point sets reconstructed by SfS and multiview stereo. For sorting out these two kinds of the point sets, a two-phased point removal is achieved: 1) pruning of stereo points based on their reliability and 2) globally-optimized carving of the surface of a visual hull by using the stereo points with graph-cuts.

## References

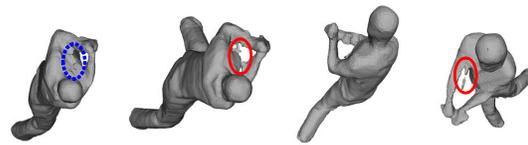
- [1] K. N. Kutulakos and S. M. Seitz, "A Theory of Shape by Space Carving," *IJCV*, Vol.38, No.3, 2000.
- [2] T. Tung, S. Nobuhara, and T. Matsuyama, "Simultaneous super-resolution and 3D video using graph-cuts," in *CVPR*, 2008.
- [3] C. Hernandez, G. Vogiatzis, and R. Cipolla, "Probabilistic visibility for multi-view stereo," *CVPR*, 2007.



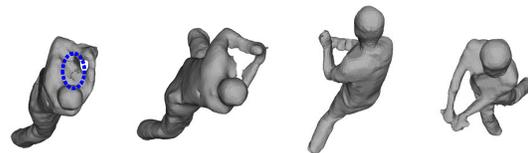
(a) Observed images



(b) Results of space carving[2]



(c) Results of shape carving[10]



(d) Results of our method

**Figure 5. Batting sequence where stereo cannot work well due to a uniform texture.**

- [4] S. Sinha and M. Pollefeys, "Multi-view reconstruction using photo-consistency and exact silhouette constraints: A maximum-flow formulation," *ICCV*, 2005.
- [5] V. Lempitsky, Y. Boykov, and D. Ivanov, "Oriented visibility for multiview reconstruction," *ECCV*, 2006.
- [6] J.-P. Pons, R. Keriven, and O. D. Faugeras, "Modelling dynamic scenes by registering multi-view image sequences," *CVPR*, 2005.
- [7] C. Strecha, R. Fransens, and L. V. Gool, "Combined depth and outlier estimation in multi-view stereo," *CVPR*, 2006.
- [8] Y. Furukawa and J. Ponce, "Accurate, Dense, and Robust MultiView Stereopsis," *PAMI*, Vol.32, No.8, pp.1362–1376, 2010.
- [9] M. Kazhdan, M. Bolitho, and H. Hoppe, "Poisson Surface Reconstruction," in *SGP*, 2005.
- [10] K. Matsuda and N. Ukita, "Direct Shape Carving: Smooth 3D Points and Normals for Surface Reconstruction," *IAPR Conference on MVA*, 2011.
- [11] Y. Boykov, O. Veksler, and R. Zabih, "MarkovRandom Fields with Efficient Approximations," *CVPR*, 1998.
- [12] Y. Boykov, O. Veksler, and R. Zabih, "Fast Approximate Energy Minimization via Graph Cuts," *PAMI*, Vol.23, No.11, pp.1222–1239, 2001.