MVA2011 IAPR Conference on Machine Vision Applications, June 13-15, 2011, Nara, JAPAN

Direct Shape Carving: Smooth 3D Points and Normals for Surface Reconstruction

Kazuki Matsuda and Norimichi Ukita
Nara Institute of Science and Technology
{kazuki-ma,ukita}@is.naist.jp

Abstract

This paper proposes a method for reconstructing a smooth and accurate 3D surface. Recent machine vision techniques can reconstruct accurate 3D points and normals of an object. The reconstructed point cloud is used for generating its 3D surface by surface reconstruction. The more accurate the point cloud, the more correct the surface becomes. For improving the surface, how to integrate existing techniques for point reconstruction is proposed. Specifically, robust and dense reconstruction with Shape-from-Silhouettes (SfS) and accurate stereo reconstruction are integrated. Unlike gradual shape shrinking by space carving, our method obtains 3D points by SfS and stereo independently and accepts the correct points reconstructed. Experimental results show the improvement by our method.

1 Introduction

Human motion capture by multiple cameras is useful for various applications such as producing 3D video contents, computer-assisted physical rehabilitation, and so on. For human motion capture, several techniques are required (e.g. camera calibration, 3D reconstruction, texture mapping, pose estimation). Among all, we focus on 3D reconstruction.

3D reconstruction from multiviews is still a hot issue in machine vision. It is briefly classified into two classes: Shape-From-Silhouettes (SfS) and Stereopsis.

In SfS, multiview silhouettes of a target object are projected to a 3D space and then 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. These false-positives are unavoidable due to the nature of SfS.

In stereopsis, image windows that match between multiviews are found in order to compute the distance to the 3D points that projected onto those windows. In principle, every point where multiview matching is established can be reconstructed. Difficulty in matching is caused in textureless or uniquely-textured regions. This difficulty results in sparse and incorrect points.

This paper proposes how to integrate the advantages of the above two schemes; silhouette constraint of SfS and photometric consistency of stereopsis. While existing algorithms refine the visual hull in an iterative manner using 3D positions and photo consistency, our new contribution is 1) direct shape carving towards the point cloud for avoiding local optima in iteration and 2) using surface normals for robust shape carving.

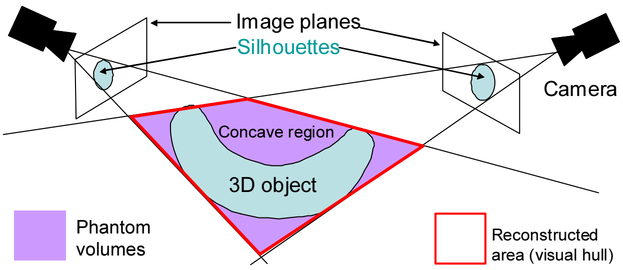


Figure 1. Shape from Silhouettes.

2 Basic Methods for 3D Surface Reconstruction

This section introduces existing algorithms for 3D surface reconstruction and their known problems, which are resolved in our proposed method.

2.1 Shape-from-Silhouettes

Figure 1 illustrates a visual hull reconstructed by SfS[1]. Even if the correct silhouettes of a target object are extracted from multiview images, the visual hull includes false-positives as well as the real shape of the object. The false-positives are called *phantom volumes*. While the amount of them is reduced as the number of the cameras grows, it is essentially impossible to remove them in the concave regions of the object.

Despite the phantom volumes, SfS is widely used for human shape reconstruction. This is because silhouette extraction is easier than stereo point correspondence in a studio, and dense and smooth surface points are obtained.

The smooth points allow us to estimate their surface normals as follows. The surface points, where their one or more neighboring points are outside the visual hull, are extracted. Then, at each surface point, an outward vector that is perpendicular to the local tangent plane defined by the surface points is regarded as the normal.

2.2 Accurate Multiview Stereo

Although early works in multiview stereo matches all points independently, recent approaches find the points on the surface that minimizes a global photometric consistency with smoothness constraints (e.g. optimized by level sets[8], and EM[9]). Novel techniques can reconstruct normals as well as 3D points; for example, patch-based multiview stereo[10].

While multiview stereo can reconstruct accurate 3D positions, it has a big disadvantage for human shape reconstruction; most cloth has less texture, which makes point correspondence difficult. This difficulty causes incorrect reconstruction and lack of the surface.

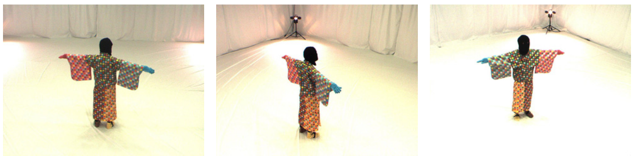
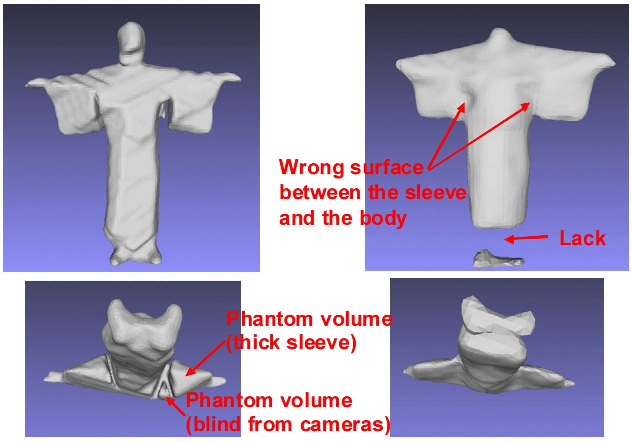


Figure 2. Images captured from multi viewpoints.



Visual hull Stereo point cloud

Figure 3. Poisson surface reconstruction from a visual hull and a point cloud using multiview images shown Fig. 2.

2.3 Existing Approaches for improving SfS and Multiview Stereo

The most popular approach for refining a visual hull is space carving[2]. The visual hull, which is an initial shape, is carved gradually until photometric consistency is satisfied between multiviews. Other constraints such as smoothness can be also optimized (e.g. using continuous local optimization by gradient descent[3], discrete global optimization by graph cuts[4, 5], and continuous global optimization[6]).

Photometric consistency becomes more powerful by employing artificial textures on a target surface. Colored patches on the cloth surface[11] can be used for human shape reconstruction.

For reconstructing a structurally-known shape such as a human shape, the shape model (e.g. articulated mesh[7]) is useful for improving accuracy and robustness to noise.

The approaches mentioned in this section have the following limitations:

- Gradual carving from a visual hull tends to fall into local optima.
- Specially-colored cloth is unavailable in natural scenarios.
- The articulated model with rigid motion of limbs is inapplicable to non-rigid loose-fitting clothing such as a skirt, floaty dress, or Japanese kimono.

2.4 Poisson Surface Reconstruction

3D surface reconstruction from 3D points is a popular problem in Computer Graphics for producing movies and so on. In contrast to simple meshing such as Marching Cubes[12], an implicit function framework is good at filling of surface holes and smooth meshing of existing points. In this paper, Poisson surface reconstruction[13] is employed. The Poisson formulation is applied to the oriented points for coping with noise. As well as the 3D position, the orientation, which is equal to the surface normal, is required in Poisson formulation.

Figure 3 shows the results of Poisson surface reconstruction from a visual hull (left) and a point cloud reconstructed by multiview stereo[10] (right) using eight-viewpoint images shown in Fig. 2. For emphasizing the limitations of each method, specially-colored clothing was used for easy reconstruction. While the visual hull produces the feasible surface with no deleted body-regions, several body-regions (e.g. feet) are missing in the one reconstructed from the stereo point cloud. This is because no valid point correspondence was obtained in these regions. From the stereo point cloud, on the other hand, thin sleeves are reconstructed correctly although phantom volumes make the sleeves thicker than the real shape.

Our goal is to acquire 3D oriented points that allow Poisson surface reconstruction to generate the accurate surface of the whole human body.

3 Direct Shape Carving using SfS and Multiview Stereo

Existing approaches in space carving gradually remove surface points in a visual hull until photometric consistency is satisfied. Photometric consistency could be satisfied before the reconstructed surface reaches the real surface.

Instead of gradual update, our method directly accepts surface points reconstructed by multiview stereo[10] and then remove the visual hull points outside the points accepted. To this end, technical issues are as follows:

- Accept only reliable oriented points reconstructed by multiview stereo, which are not inconsistent with the visual hull.
- Find and remove visual hull points that interrupt the view lines from cameras to the accepted stereo points.

First of all, SfS and multiview stereo are applied to images independently. We rely on points reconstructed by multiview stereo, except the ones whose normals are inconsistent with those of the visual hull. This is because:

- accuracy of the normals gets much lower where the reconstructed points are sparse (while dense points are obtained in every surface by SfS, stereo might reconstruct sparse points in textureless regions) and
- inconsistent normals, which might be observed where stereo points are reconstructed sparsely near the surface of the visual hull, make it difficult to apply the Poisson formula.

Based on the criteria above, a stereo point is removed if 1) it is close to the surface of the visual hull and 2)

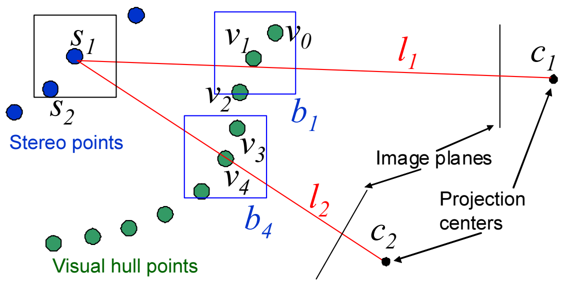
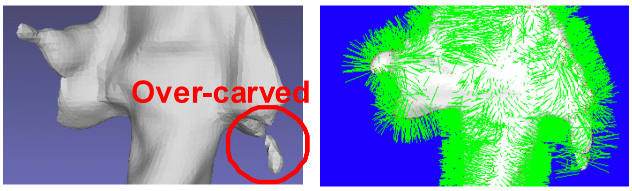


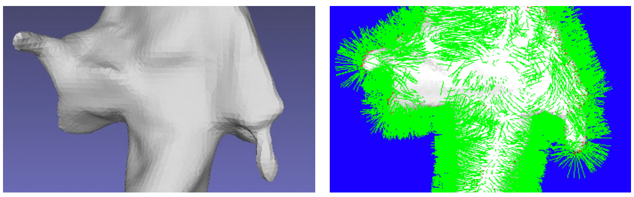
Figure 4. Shape carving by testing ray intersection with a bounding box.

Figure 5. Images captured from multi viewpoints.



Surface Surface and normals

Figure 6. Poisson surface reconstruction from all stereo points and the remaining visual hull points. Green lines indicate normals.



Surface Surface and normals

Figure 7. Poisson surface reconstruction using our method.

its normal is significantly different from the normal of the nearest visual hull surface. Specifically, if 1) the distance to the nearest visual hull surface is shorter than $w_l l$, where l and w_l denote the side length of a voxel and a constant, respectively, and 2) the angle between the normals of the stereo point of interest and the nearest visual surface is larger than a threshold, θ_a, the stereo point of interest is removed. In our experiments, $w_l = 8$ and $\theta_a = 30$ degrees.

Like ray tracing, each of the remaining stereo points is projected onto the image planes in which the point is extracted. If the ray hits one or more points reconstructed by SfS, these points are removed as phantom volumes. Actually, the bounding box around each point is prepared for this intersection test[14]; if the ray crosses the box, its respective visual hull point is removed.

For this intersection test, the size of the bounding box is critical. If the size is smaller/larger, visual hull points that must be removed/remained are remained/removed incorrectly. The size is determined in accordance with the distance between the stereo point of interest and its nearest stereo point. Figure 4 illustrates this intersection test. Given a stereo point, s_1 and its nearest point, s_2, the size of the bounding box is equal to the cube whose center is s_1 and that passes s_2. This bounding box is located in every visual hull point (e.g. box b_1 is located in point v_1 in Fig. 4). 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 l_1) passes through b_1, v_1 is removed. v_0 and v_2 are also removed because l_1 passes through their bonding boxes. Similarly, v_3 and v_4 are also removed because their bounding boxes are on the way of l_2 that is from s_1 to c_2.

Finally, with the remaining stereo points and visual hull points, the 3D surface is acquired by Poisson surface reconstruction[13].

As mentioned above, our method consists of pruning stereo points and carving visual hull points. Carving is essentially required as in the way of space carving. On the other hand, one might use all stereo points with no pruning because careful optimization in multiview stereo obtains correct 3D points.

Figure 6, which shows the surface reconstructed from images in Fig. 5, is a typical example that proves the demand of the pruning. The left-hand image is the surface reconstructed from all stereo points, the remaining visual hull points, and their normals, which are shown in the right-hand image. Although the points of the left hand were reconstructed, it was over-carved. This was caused due to sparse scattered normals in that region; see the right-hand image. These scattered normals make the Poisson formulation difficult to solve.

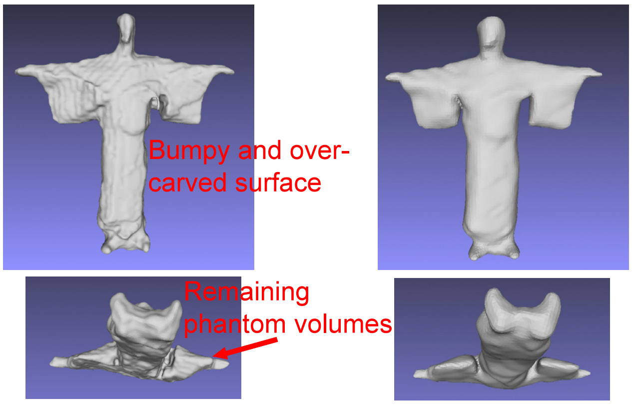
In contrast to Fig. 6, the result of our method (Fig. 7) looks better because of smooth normals even in the hand region. This is why pruning scattered normals is required before surface reconstruction.

To prove the performance of our method in carving phantom volumes, the surfaces reconstructed by space carving with graph-cuts[5] and our method were compared. The results are shown in Fig. 8: space carving with graph-cuts (left) and our method (right). The phantom volumes remained in the result of space carving, while our method could successfully carve them. It can be also seen that the surface of space carving was very bumpy. This might be caused due to local optima in carving.

4 Experiments with Image Sequences

Our method was applied to multiview image sequences for validating the effectiveness of our method.

Figures 9 – 14 show observed images and the results of surface reconstruction from them. For comparison, the results by space carving[5] are also shown. All silhouettes and parameters in Poisson surface reconstruction were same between two methods (i.e. space carving and our method). All the results were obtained from eight cameras. The cameras were located around a subject. If a camera was located so that it looked down on the subject from the above, it could reduce phantom volumes efficiently by SfS, especially those surrounded by the arms. But no camera was located above the subject for examining the performance of carving them under stringent conditions. It can be



Carving with graph-cuts Our method

Figure 8. Poisson surface reconstruction from the results of space carving and our method using multiview images shown Fig. 2.

Figure 9. Dance sequence: observed images.

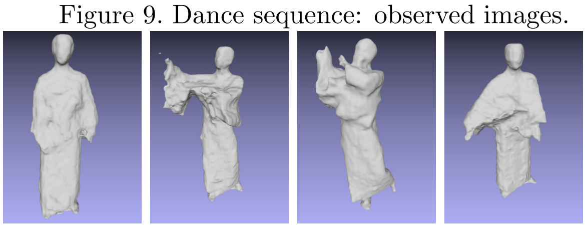


Figure 10. Dance sequence: space carving[5].

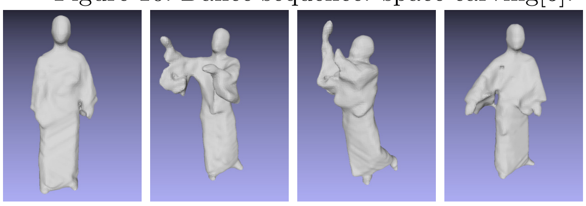


Figure 11. Dance sequence: our method.

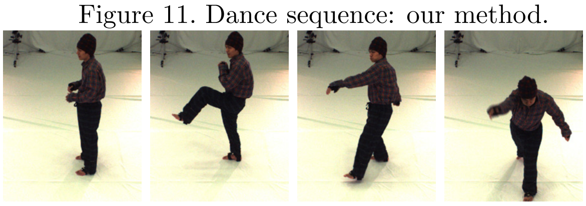


Figure 12. Throwing sequence: observed images.

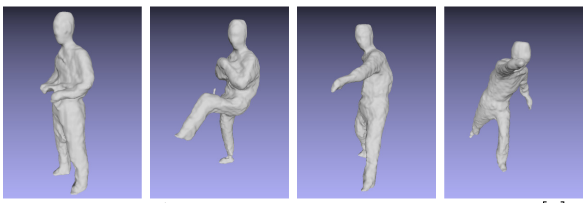


Figure 13. Throwing sequence: space carving[5].

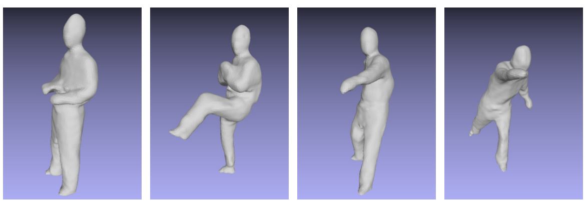


Figure 14. Throwing sequence: our method.

seen that our method could reconstruct smooth surfaces with less phantom volumes in contrast to space carving.

5 Concluding Remarks

This paper proposed 3D oriented point reconstruction from multiviews. The method employs two kinds of point sets reconstructed by SfS and optimized multiview stereo. For sorting out these two kinds of the point sets, a two-phased point removal is achieved: 1) pruning based on irregularity of the normals of the points and 2) carving along the rays from the optimized stereo points to the cameras that observe the points of interest.

Future work includes reoptimization around boundaries between the points reconstructed by SfS and multiview stereo for a more smooth surface.

The source codes of PMVS[10], Poisson surface reconstruction[13], and space carving with graph-cuts[5] were given by Y. Furukawa, M. Kazhdan, and S. Nobuhara, respectively.

References

[1] G. Cheung, T. Kanade, J. Bouguet, and M. Holler, "A real time system for robust 3D voxel reconstruction of human motions," *CVPR*, 2000.

[2] K. N. Kutulakos and S. M. Seitz, "A Theory of Shape by Space Carving," *IJCV*, Vol.38, No.3, 2000.

[3] C. H. Esteban and F. Schmitt, "Silhouette and stereo fusion for 3D object modeling," *CVIU*, Vol.96, No.3, 2004.

[4] G. Vogiatzis, P. H. Torr, and R. Cipolla, "Multi-view stereo via volumetric graph-cuts," *CVPR*, 2005.

[5] T. Tung, S. Nobuhara, and T. Matsuyama, "Simultaneous super-resolution and 3D video using graph-cuts," in *CVPR*, 2008.

[6] K. Kolev, M. Klodt, T. Brox, D. Cremers, "Continuous Global Optimization in Multiview 3D Reconstruction," *IJCV*, Vol.84, No.1, pp.80–96, 2009.

[7] E. de Aguiar, C. Stoll, C. Theobalt, N. Ahmed, H.-P. Seidel, S. Thrun, "Performance capture from sparse multi-view video," *ACM SIGGRAPH*, 2008.

[8] J.-P. Pons, R. Keriven, and O. D. Faugeras, "Modelling dynamic scenes by registering multi-view image sequences," *CVPR*, 2005.

[9] C. Strecha, R. Fransens, and L. V. Gool, "Combined depth and outlier estimation in multi-view stereo," *CVPR*, 2006.

[10] Y. Furukawa and J. Ponce, "Accurate, Dense, and Robust MultiView Stereopsis," *PAMI*, Vol.32, No.8, pp.1362–1376, 2010.

[11] R. White, K. Crane, and D. Forsyth, "Capturing and Animating Occluded Cloth," ACM *TOG (SIGGRAPH)*, Vol.26, No.3, 2007.

[12] W. E. Lorensen and H. E. Cline, "Marching Cubes: A high resolution 3D surface construction algorithm," *Computer Graphics*, Vol.21, No.4, 1987.

[13] M. Kazhdan, M. Bolitho, and H. Hoppe, "Poisson Surface Reconstruction," in *SGP*, 2005.

[14] T. Moller and B. Trumbore, "Fast, Minimum Storage Ray-Triangle Intersection," *Journal of Graphic Tools*, Vol.2, No.1, pp.21–28, 1997.