

3D Reconstruction by Combining Shape from Silhouette with Stereo

Huei-Yung Lin and Jing-Ren Wu
Department of Electrical Engineering
National Chung Cheng University
Chia-Yi 621, Taiwan, R.O.C.
lin@ee.ccu.edu.tw, jr@image.ee.ccu.edu.tw

Abstract

In this paper we propose a 3D reconstruction algorithm by combining shape from silhouette with stereo. Visual hull of the object is first derived from multi-view silhouette images. Pairwise stereo matching for shape refinement is then accomplished using the best viewable images. Based on the reduced correspondence searching range constrained by contact points and bounding edges, significant improvement of visual hull is possible even if the number of cameras is limited. Experimental results are presented for both synthetic data and real scene images.

1 Introduction

3D model reconstruction of a real scene is an important and active research topic in computer vision. It has many applications ranging from industrial inspection and reverse engineering to computer graphics and multimedia. The objective is usually to recover the geometric (and possibly the photometric) information of the scene using intensity images recorded by a camera. The acquired 3D measurements can then be used to generate a computer model of the scene.

In the past few decades many 3D reconstruction algorithms based on different visual cues, such as stereo, motion, shading, silhouette, texture, have been proposed [12]. For a given real object, however, not all of the above methods are capable of creating a complete 3D model without acquisition, registration and data fusion of multiple range images [14]. For example, stereo vision or shape from shading can only provide the so-called 2.5D range data from a single viewpoint. Multiple image captures from different viewing directions are mandatory for the reconstruction of a complete 3D model. In addition to the 3D alignment between different range data sets, the reconstruction of a dynamic

scene is also not possible without simultaneous multi-view 3D recovery.

Shape from silhouette, on the other hand, reconstruct a 3D model using the object silhouettes acquired from the surrounding cameras [7]. Since the complete 3D model is recovered by the intersection of all silhouette cones back-projected from the camera centers, it can be implemented very easily and is also suitable for the reconstruction of moving objects. One major drawback of this approach is that the resulting *visual hull* is generally not a good geometric approximation of the observed shape. It might be even worse if the number of cameras is reduced or the object consists of apparent concave surface shape. Thus, a number of methods have been proposed to improve shape from silhouette with additional constraints [6, 2, 1, 11].

Our goal is to refine shape from silhouette with stereo to extract more precise geometric models. Several approaches for 3D reconstruction by exploiting the advantages of combining these two techniques have been investigated in recent research. Li et al. construct a polyhedral visual hull from silhouettes as an initial estimate, and it is then used to restrict the disparity searching range for stereo [8]. Esteban et al. generate an octree-based coarse model from visual hull followed by a multi-stereo carving technique for refinement [5]. Cheung et al. assume the object is under rigid motion and improve shape from silhouette by registering and refining the visual hulls across time [2].

In this paper, we present a 3D reconstruction algorithm using the visual hull derived from the object silhouettes with the refinement of pairwise stereo matching. Different from the previous approaches, our 3D surface is refined based on the best viewable stereo image pair. Given the cameras with fixed positions and orientations, the 3D reconstruction result is more precise for certain viewpoints over others. It is therefore suitable for stereoscopic image synthesis with predetermined primary viewpoints.

2 Shape from Silhouette

Suppose an object is placed in front of multiple cameras and its silhouettes are identified in all images. For any given camera, the viewing cone is defined by a set of viewing edges – the rays connecting the silhouette boundary and the center of projection. The intersection of all viewing cones is called the visual hull of the object. It is clear that the object surface is bounded by its visual hull. By identifying the bounding edges (i.e. the line segments of the viewing edges contained in all silhouettes) of the visual hull, a rough 3D model of the object can be approximated by a single surface representation.

Consider a set of silhouettes $S_i, i = 1, 2, \dots, N$, with u_i^j the boundary points of the silhouettes. The viewing edge r_i^j is obtained by back-projecting u_i^j from the camera center C_i . Suppose the projection of a 3D scene point onto the k -th camera is given by $\Pi_k(\cdot)$, then the bounding edge E_i^j is a subset of r_i^j which satisfies the condition $\Pi_k(E_i^j) \subset S_k$ for all $k = 1, 2, \dots, N$. It should be noted that the bounding edge could consist of several disjointed line segments.

One way to derive the bounding edge E_i^j is to form the line segments $\Pi_k(r_i^j) \cap S_k$ in all images, and have them back-projected to the 3-space for intersection with the viewing edge r_i^j as described in [2]. This approach, however, does not efficiently reduce the bounding edge searching range. In this work, the best camera viewpoint C_k with respect to the i -th camera is first determined by the relationship between r_i^j and all other centers of projection. The intersection of the corresponding epipolar line in the k -th image and the silhouette S_k is then used to derive a base 3D line segment L_i^k . The bounding edge is then given by the part of L_i^k satisfying $\Pi_j(L_i^k) \cap S_j, j \neq i, k$, as illustrated in Figure 1.

3 Refining Visual Hull with Stereo

The fundamental principle of stereo is to infer the depth information based on the point correspondence obtained in the image pair. In addition to a robust metric for template matching, how to restrict the searching range is also a key to success of stereo matching algorithms [10]. In this paper, we emphasize on reducing the stereo searching range based on the result provided by the initial visual hull. Similar to the idea of registration assisted stereo matching proposed by Lin et al. [9], the spheres centered at the “contact points” of the object surface with predetermined radii are used to restrict the correspondence searching ranges.

One important property of bounding edge is that it touches the object at at least one point, which is called a

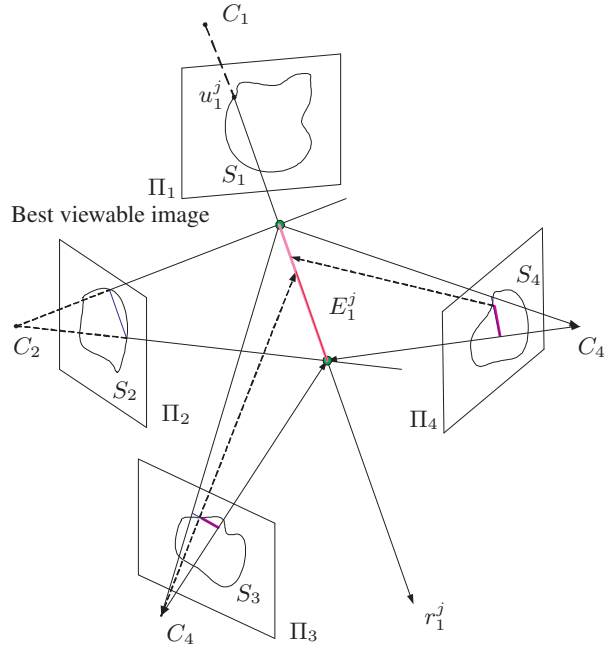


Figure 1. Interpretation and derivation of the bounding edge.

“contact point” of the surface. Since any bounding edge delimits the object, the containing edge points closer to a contact point generally implies that they are better approximations to the real surface. Based on this observation, De Roeck et al. [3] use visual hull for plane sweeping and improve stereo matching with the surface “anchor points”. Franco et al. [4] use the relationship between contact points and viewing edges to approximate the so-called “visual shapes” under smoothness assumption of the visible surface.

Our approach utilizes the distance between the contact point and the endpoints of the corresponding bounding edge to define a sphere centered at the contact point. The sphere is then used to constrain the depth range associated with the bounding edge for correspondence matching. It is clear that the size of the sphere depends on the length of the bounding edge as well as the location of the contact point on the bounding edge. When the number of viewpoints increases for shape from silhouette, the number of bounding edges becomes larger while the average length becomes shorter in general. Consequently, the sizes of the spheres and thus the searching ranges are inversely proportional to the number of cameras. The performance increases more significantly as more and more cameras are added to the system.

To refine a visual hull with the above stereo matching algorithm, the best viewable stereo image pair has

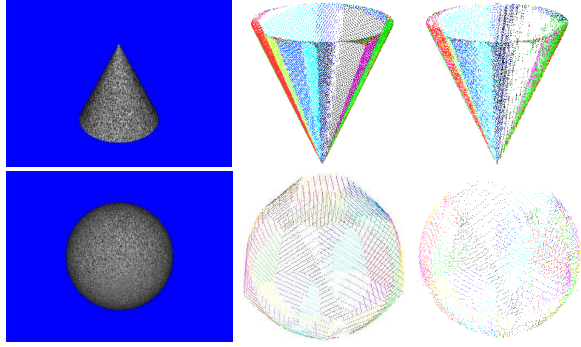


Figure 2. Simulation results of synthetic data set used for error analysis. From the left to the right: one of the eight rendered images, 3D reconstruction by shape from silhouette, stereo refinement with the proposed algorithm.

to be identified. Consider the bounding edge E_m^j originated from the m -th image, the corresponding image for stereo matching is selected by the one with the smallest angle between its optical axis and the viewing edge r_m^j . Mathematically, the stereo image pair (I_m, I_n) is determined by

$$\arg \max_{m,n} \cos(r_m^j, o_n)$$

where o_n represents the optical axis of the camera C_n . Since the viewing edge r_m^j is not fixed for different boundary point u_m^j of the silhouette S_m , the corresponding image for stereo matching varies with the position of u_m^j .

Once the best viewable stereo image pair is obtained, the contact point on any bounding edge can be identified by template matching and the epipolar constraint. A sphere can then be created based on the length of the bounding edge, and the depth range is estimated within the sphere radius. Note that, since the contact points are the points contained in the visual hull and closest to the true object surface, the depth range for correspondence searching can be further restricted to the same side with respect to the bounding edge.

The presented algorithm for stereo-based visual hull refinement can be summarized as follows. The bounding edges generated by shape from silhouette are used to create the spheres centered at the containing contact points. The circle projections of the spheres onto the best viewable stereo image pairs are then used to restrict the correspondence searching range. Thus, based on the robust but sparse bounding edges derived from silhouettes, a dense 3D model is reconstructed by stereo with highly constrained correspondence searching regions.

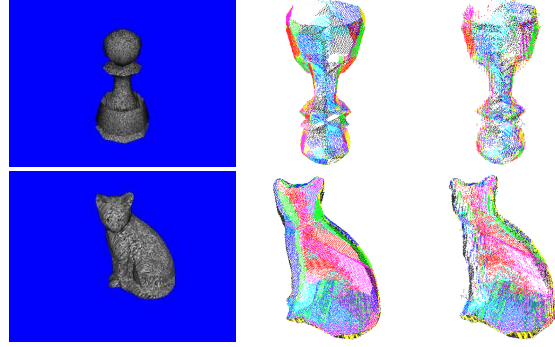


Figure 3. 3D reconstruction of a pawn and a cat object from synthetic data set.

4 Experimental Results

Simulation with synthetic data sets is carried out first to demonstrate the validity of our algorithm. Two textured 3D computer models, a cone and a sphere, are generated and rendered using VTK (The Visualization Toolkit) with eight surrounding virtual cameras. Figure 2 shows one of the captured images, the results of shape from silhouette and the proposed method, respectively, for both objects. Different color in the results represents 3D reconstruction from different camera. The error analysis in terms of mean absolute deviation (MAD) is tabulated in Table 1. The base radius and height of the cone are 4 mm and 8 mm, respectively. The radius of the sphere is 5 mm. Figure 3 shows the visual hulls and stereo refinements of more complex objects.

Our experimental setup for 3D model reconstruction includes eight cameras surrounding the object and facing downward about 45 degrees. The intrinsic and extrinsic camera parameters are calibrated using Tsai's method [13]. Figure 4 shows the foreground images acquired from all viewpoints. The 3D models reconstructed using shape from silhouette and the proposed method are illustrated in Figures 5(a) and 5(b), respectively. The head and other body parts of the object appear more smooth by the stereo refinement. Depending on the object size, the processing time with the proposed stereo refinement generally increases about 20% – 60% compared to the computation of shape from silhouette alone.

Table 1. MAD of the virtual objects.

Object model	Cone	Sphere
Shape from silhouette	0.13 (mm)	0.10 (mm)
The proposed method	0.08 (mm)	0.09 (mm)

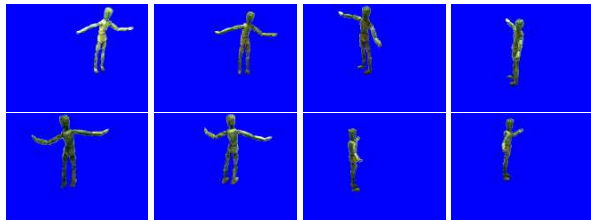


Figure 4. Foreground images of a mannequin object acquired from all cameras.

5 Conclusion

In this paper we have presented a 3D reconstruction algorithm to refine shape from silhouette with stereo. Different from the previous approaches, our 3D surface refinement is based on the best viewable stereo image pair. Thus, the improvement over shape from silhouette is significant even if the numbers of viewpoints and bounding edges are limited. Currently the surface regions between different stereo image pairs is not considered explicitly, stereo algorithms dealing with large baselines will be investigated for shape refinement in the future work.

Acknowledgments

The support of this work in part by the National Science Council of Taiwan, R.O.C, under Grant NSC-96-2221-E-194-016-MY2 is gratefully acknowledged.

References

- [1] E. Boyer and J. Franco. A hybrid approach for computing visual hulls of complex objects. In *IEEE Computer Vision and Pattern Recognition*, pages I: 695–701, 2003.
- [2] K. Cheung, S. Baker, and T. Kanade. Shape-from-silhouette across time part i: Theory and algorithms. *International Journal of Computer Vision*, 62(3):221–247, May 2005.
- [3] S. de Roeck, N. Cornelis, and L. Van Gool. Augmenting fast stereo with silhouette constraints for dynamic 3d capture. In *International Conference on Pattern Recognition*, pages I: 131–134, 2006.
- [4] J. Franco, M. Lapierre, and E. Boyer. Visual shapes of silhouette sets. In *3D Data Processing, Visualization and Transmission*, pages 397–404, 2006.
- [5] C. Hernandez Esteban and F. Schmitt. Silhouette and stereo fusion for 3d object modeling. *Computer Vision and Image Understanding*, 96(3):367–392, December 2004.



(a) Shape from silhouette.



(b) Stereo refinement using the proposed algorithm.

Figure 5. 3D reconstruction results of a real mannequin object.

- [6] K. N. Kutulakos and S. M. Seitz. A theory of shape by space carving. *International Journal of Computer Vision*, 38(3):199–218, 2000.
- [7] A. Laurentini. The visual hull concept for silhouette-based image understanding. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 16(2):150–162, February 1994.
- [8] M. Li, H. Schirmacher, M. Magnor, and H.-P. Siedel. Combining stereo and visual hull information for on-line reconstruction and rendering of dynamic scenes. *Multimedia Signal Processing, 2002 IEEE Workshop on*, pages 9–12, 9-11 Dec. 2002.
- [9] H. Lin, S. Liang, and J. Wu. 3d shape recovery with registration assisted stereo matching. In *Iberian Conference on Pattern Recognition and Image Analysis*, pages II: 596–603, 2007.
- [10] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1-3):7–42, April 2002.
- [11] S. Seitz and C. Dyer. Photorealistic scene reconstruction by voxel coloring. *International Journal of Computer Vision*, 35(2):151–173, November 1999.
- [12] E. Trucco and A. Verri. *Introductory Techniques for 3-D Computer Vision*. Prentice Hall, 1998.
- [13] R. Tsai. A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Trans. Robotics and Automation*, 3(4):323–344, 1987.
- [14] M. Weber, A. Blake, and R. Cipolla. Towards a complete dense geometric and photometric reconstruction under varying pose and illumination. *Image and Vision Computing*, 22(10):787–793, September 2004.