International Journal of Innovative
Computing, Information and Control
Volume 10, Number 5, October 2014

ICIC International ©2014 ISSN 1349-4198
pp. 1915–1921

A NOVEL 3D RECONSTRUCTION APPROACH

Huei-Yung Lin[1,2], Chin-Chen Chang[3,*] and Jing-Ren Wu[1,2]

[1]Department of Electrical Engineering
[2]Advanced Institute of Manufacturing with High-Tech Innovations
National Chung Cheng University
No. 168, Sec. 1, University Rd., Min-Hsiung Township, Chia-yi County 62102, Taiwan

[3]Department of Computer Science and Information Engineering
National United University
No. 1, Lienda, Miaoli 36003, Taiwan
[*]Corresponding author: ccchang@nuu.edu.tw

Received December 2013; revised April 2014

Abstract. *3D model reconstruction of a real scene is an important and active research topic in computer vision. In this paper, we propose a novel 3D reconstruction algorithm by combining shape from silhouette with stereo. We first construct a visual hull of the object from multi-view silhouette images. Then we accomplish pairwise stereo matching for shape refinement using the best viewable images. Based on the reduced correspondence searching range constrained by contact points and bounding edges, visual hull is significantly improved even if the number of cameras is limited. Results show that our approach performs well for both synthetic data and real scene images.*
Keywords: 3D reconstruction, Shape from silhouette, Stereo, Visual hull

1. **Introduction.** In computer vision, reconstructing a 3D model of a real scene is very important. There are many applications ranging from industrial inspection to computer graphics and multimedia. The objective of 3D model reconstruction for a scene is usually to recover the geometric information of the scene using intensity images recorded by a camera. In the past decades, many 3D reconstruction algorithms [16] have been proposed. These approaches are mostly based on different visual cues, such as stereo, motion, shading, silhouette, and texture. For a given real object, however, not all of the above methods can create a complete 3D model without acquisition, registration and data fusion of multiple range images [18]. For example, stereo vision or shape from shading can only provide the 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 multiview 3D recovery.

On the other hand, the technique of shape from silhouette reconstructs a 3D model using the object silhouettes acquired from the surrounding cameras [11]. The complete 3D model is recovered by the intersection of all silhouette cones back-projected from the camera centers. This technique can be easily implemented and is also suitable for the reconstruction of moving objects. One major drawback of this technique 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, several approaches have been proposed to improve shape from silhouette with additional constraints [2,3,10,15].

In this paper, we present a novel 3D reconstruction algorithm which combines shape from silhouette with stereo. The contributions of the proposed approach are as follows: a) we use the visual hull derived from the object silhouettes with the refinement of pairwise stereo matching; b) 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; c) our approach is suitable for stereoscopic image synthesis with predetermined primary viewpoints.

The rest of the paper is organized as follows. Section 2 reviews related works. Section 3 presents the proposed approach. Results are given in Section 4. Finally, conclusions are discussed in Section 5.

2. Related Works.

Several approaches for 3D reconstruction by exploiting the advantages of combining these two techniques have been investigated. Li et al. [12] constructed a polyhedral visual hull from silhouettes as an initial estimate. It is then used to restrict the disparity searching range for stereo. Esteban and Schmitt [7] generated an octree-based coarse model from visual hull followed by a multi-stereo carving technique for refinement. Cheung et al. [3] assume that the object is under rigid motion and improve shape from silhouette by registering and refining the visual hulls across time.

Zeng et al. [19] presented an approach for large-scale 3D scene reconstruction and image grouping algorithm for unordered wide-baseline photos. Their approach can solve the large-scale 3D reconstruction problem. Kolev et al. [9] proposed a global optimization approach to the field of multiview 3D reconstruction. They presented an approach to cast the 3D shape reconstruction problem as minimizing a spatially continuous convex functional. Beall et al. [1] presented an approach for 3D reconstruction of underwater structures. They constructed the models from synchronized videos. Their approach can obtain a highly accurate sparse 3D reconstruction of underwater structures. Geiger et al. [6] proposed an approach to construct 3D maps from stereo sequences in real-time. Their method can achieve certain accuracy. Izadi et al. [8] presented a reconstruction system called KinectFusion. This system takes depth data from a Kinect camera and creates a 3D model in real-time.

3. The Proposed Approach.

3.1. Shape from silhouette.

Let an object be placed in front of multiple cameras and its silhouettes are identified in all images. For a given camera, the viewing cone is defined by a set of viewing edges. The rays connect the silhouette boundary and the center of projection. The intersection of all viewing cones is called the visual hull of the object. Hence, the object surface is bounded by its visual hull. By identifying the bounding edges 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, $i = 1, 2, \ldots, N$, 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. Supposing the projection of a 3D scene point onto the k-th camera is given by $\prod_k(\cdot)$, then the bounding edge E_i^j is a subset of r_i^j satisfying the condition $\prod_k(E_i^j) \subset S_k$ for all $k = 1, 2, \ldots, N$. The bounding edge consists of several disjointed line segments.

One way to derive the bounding edge E_i^j is to form the line segments $\prod_k(E_i^j) \cap S_k$ in all images, and have them back-projected to the 3-space for intersection with the viewing edge r_i^j [3]. This approach, however, does not efficiently reduce the bounding edge searching range. In our approach, the best camera viewpoint C_k with respect to i-th camera is first determined by the relationship between r_i^j and all other centers of projection. The

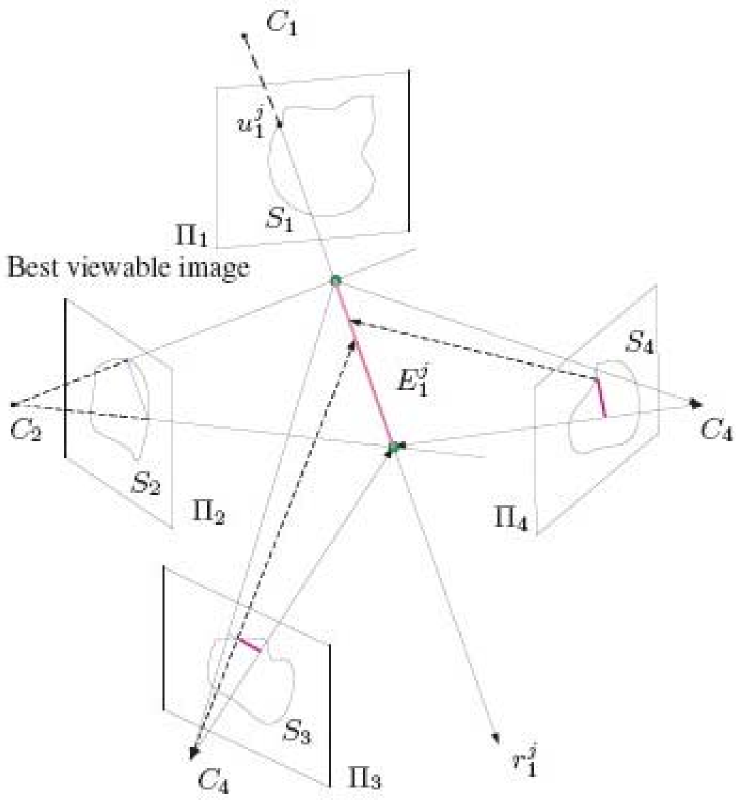

FIGURE 1. Interpretation and derivation of the bounding edge

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 $\prod_j (L_i^k) \cap S_j$, $j \neq i, k$, as shown in Figure 1.

3.2. **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, restricting the search range is also a key to success of stereo matching algorithms [14]. In our approach, 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 [13], 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 a bounding edge is that it touches the object at least one point, which is called a "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 the observation, De Roeck et al. [4] used visual hull for plane sweeping and improved stereo matching with the surface "anchor points". Franco et al. [5] used 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. 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 generally. 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 to be identified. Considering 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. 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.

Our approach 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.

4. **Results.** Several experiments were conducted to evaluate the effectiveness of the proposed approach. The proposed algorithm was implemented in C++ running on a PC with Pentium-4 2.8GHz dual core and 1 GB of system memory.

First, simulation with synthetic data sets is performed to demonstrate the validity of our approach. 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 our approach, respectively, for both objects. Different color in the results represents 3D reconstruction from different camera. The base radius and height of the cone are 4 mm and 8 mm, respectively. The radius of the sphere is 5 mm. The error analysis in terms of mean absolute deviation (MAD) and the computation cost are shown in Table 1 and Table 2, respectively. The MAD of our approach is smaller than that of the previous method. Moreover, the computation cost of our approach is larger than that of the previous method. Figure 3 shows the visual hulls and stereo refinements of more complex objects. After improving corresponding stereo image pairs, our approach can effectively eliminate the cutting angles generated by multi-view silhouettes. Hence, three-dimensional points are more close to the surface of the object.

In the following experiment, the 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 the method [17]. Figure 4

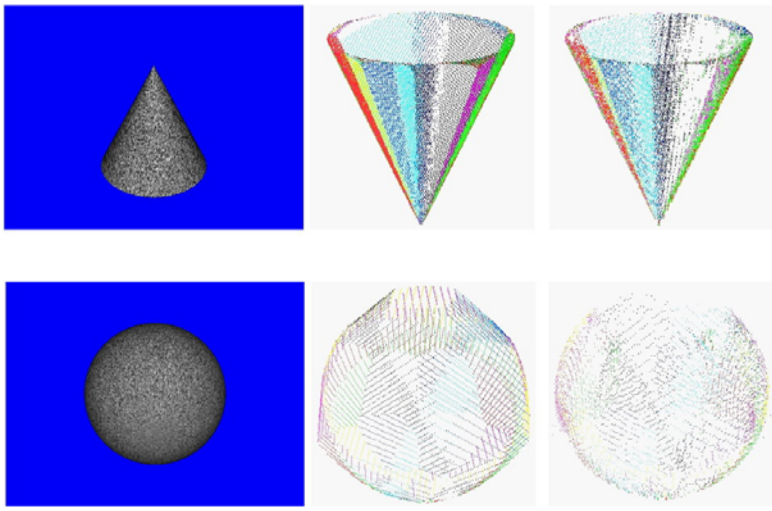

FIGURE 2. Simulation results of synthetic data set used for error analysis. From the left to the right: one of the eight rendered images, 3D reconstruction by shape from silhouette, and stereo refinement with the proposed approach.

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)

TABLE 2. Computational time of the virtual objects

Object model	Cone	Sphere
Shape from silhouette	27.2 (sec.)	7.00 (sec.)
The proposed method	33.39 (sec.)	10.56 (sec.)

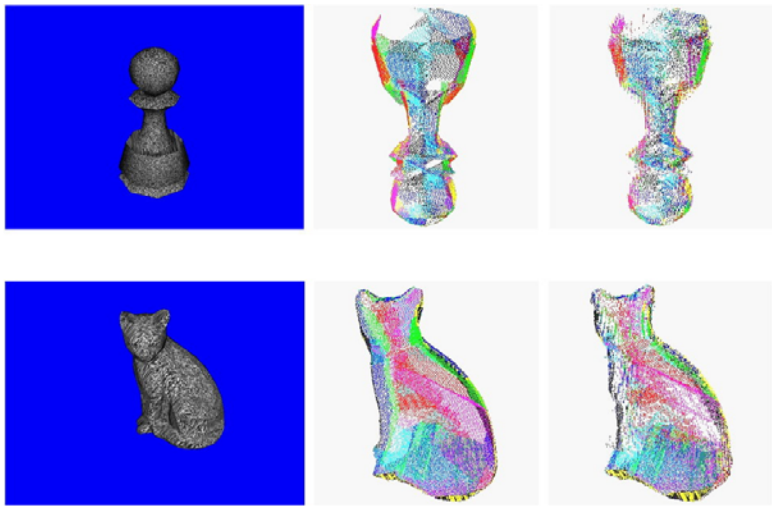

FIGURE 3. 3D reconstruction of a pawn and a cat object from synthetic data set

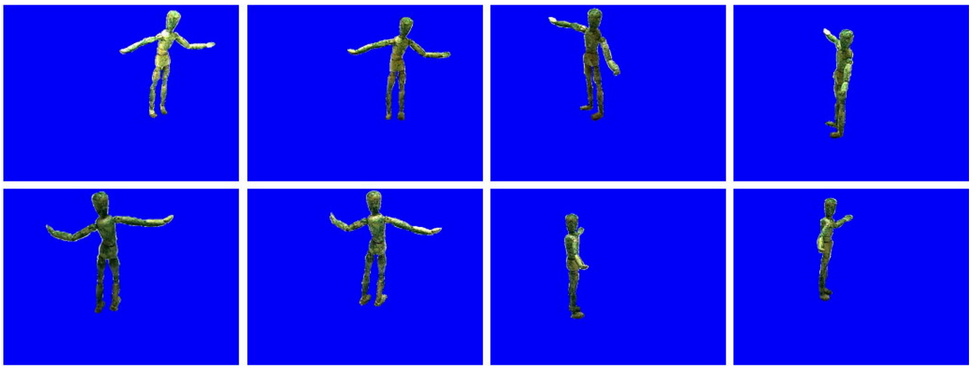

Figure 4. Foreground images of a mannequin object acquired from all cameras

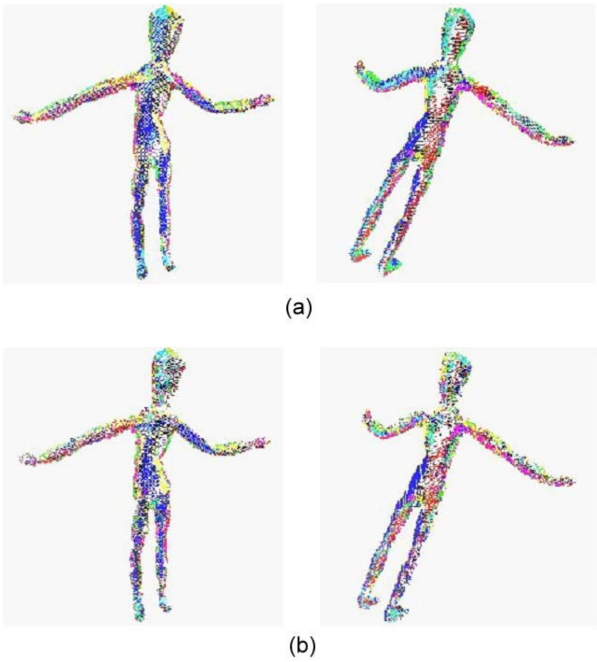

Figure 5. 3D reconstruction results of a real mannequin object: (a) shape from silhouette; (b) stereo refinement using the proposed approach

shows the foreground images acquired from all viewpoints. The 3D models reconstructed using shape from silhouette and our approach are illustrated in Figures 5(a) and 5(b), respectively. The head and other body parts of the object appear more smoothly 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.

5. **Conclusions.** We have presented a 3D reconstruction algorithm to refine shape from silhouette with stereo. Different from the previous approaches, the proposed 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 are not considered explicitly. Stereo algorithms dealing with large baselines will be investigated for shape refinement.

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

REFERENCES

[1] C. Beall, B. J. Lawrence, V. Ila and F. Dellaert, 3D reconstruction of underwater structures, *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp.18-22, 2010.

[2] E. Boyer and J. Franco, A hybrid approach for computing visual hulls of complex objects, *Proc. of IEEE Computer Vision and Pattern Recognition*, pp.695-701, 2003.

[3] K. Cheung, S. Baker and T. Kanade, Shape-from silhouette across time part I: Theory and algorithms, *International Journal of Computer Vision*, vol.62, no.3, pp.221-247, 2005.

[4] S. De Roeck, N. Cornelis and L. Van Gool, Augmenting fast stereo with silhouette constraints for dynamic 3D capture, *Proc. of International Conference on Pattern Recognition*, pp.131-134, 2006.

[5] J. Franco, M. Lapierre and E. Boyer, Visual shapes of silhouette sets, *Proc. of the 3rd International Symposium on 3D Data Processing, Visualization and Transmission*, pp.397-404, 2006.

[6] A. Geiger, J. Ziegler and C. Stiller, StereoScan: Dense 3D reconstruction in real-time, *Proc. of IEEE Intelligent Vehicles Symposium*, 2011.

[7] C. H. Esteban and F. Schmitt, Silhouette and stereo fusion for 3D object modeling, *Computer Vision and Image Understanding*, vol.96, no.3, pp.367-392, 2004.

[8] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. A. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. J. Davison and A. W. Fitzgibbon, KinectFusion: Real-time 3D reconstruction and interaction using a moving depth camera, *Proc. of ACM Symposium on User Interface Software and Technology*, pp.559-568, 2011.

[9] K. Kolev, M. Klodt, T. Brox and D. Cremers, Continuous global optimization in multiview 3D reconstruction, *International Journal of Computer Vision*, vol.84, no.1, pp.80-96, 2009.

[10] K. N. Kutulakos and S. M. Seitz, A theory of shape by space carving, *International Journal of Computer Vision*, vol.38, no.3, pp.199-218, 2000.

[11] A. Laurentini, The visual hull concept for silhouette based image understanding, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.16, no.2, pp.150-162, 1994.

[12] M. Li, H. Schirmacher, M. Magnor and H. P. Siedel, Combining stereo and visual hull information for online reconstruction and rendering of dynamic scenes, *Proc. of IEEE Workshop on Multimedia Signal Processing*, pp.9-12, 2002.

[13] H. Lin, S. Liang and J. Wu, 3D shape recovery with registration assisted stereo matching, *Proc. of Iberian Conference on Pattern Recognition and Image Analysis*, pp.596-603, 2007.

[14] D. Scharstein and R. Szeliski, A taxonomy and evaluation of dense two-frame stereo correspondence algorithms, *International Journal of Computer Vision*, vol.47, no.1-3, pp.7-42, 2002.

[15] S. Seitz and C. Dyer, Photorealistic scene reconstruction by voxel coloring, *International Journal of Computer Vision*, vol.35, no.2, pp.151-173, 1999.

[16] E. Trucco and A. Verri, *Introductory Techniques for 3-D Computer Vision*, Prentice Hall, 1998.

[17] R. Tsai, A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses, *IEEE Transactions on Robotics and Automation*, vol.3, no.4, pp.323-344, 1987.

[18] M. Weber, A. Blake and R. Cipolla, Towards a complete dense geometric and photometric reconstruction under varying pose and illumination, *Image and Vision Computing*, vol.22, no.10, pp.787-793, 2004.

[19] X. Zeng, Q. Wang and J. Xu, MAP model for large-scale 3D reconstruction and coarse matching for unordered wide-baseline photos, *Proc. of the British Machine Vision Conference*, 2008.