

# A Integrated Depth Fusion Algorithm for Multi-View Stereo

Yongjian Xi · Ye Duan

**Abstract** In this paper, we propose a new integrated depth fusion algorithm for multi-view stereo. The experiments on the benchmark datasets show that our algorithm can obtain high quality reconstruction results that are comparable with the state-of-art methods, with considerable less computational time and complexity.

**Keywords** multi-view stereo · graph cut

## 1 Introduction

Despite significant advancement in interactive shape modeling, creating complex high quality realistic looking 3D models from scratch is still a very challenging task. Recent advancement in 3D shape acquisition systems such as laser range scanners and encoded light projecting system have made directly 3D data acquisition feasible [1]. These active 3D acquisition systems however remain expensive. Meanwhile, the price of digital cameras and digital video cameras keeps decreasing while the quality is improving every day, partially due to the intense competition in the huge consumer market. Furthermore, huge amounts of images and videos are added in internet sites such as Google, etc every day, a lot of which could be used for multiview image-based 3D shape reconstruction [2].

To date, there have been a lot of researches conducted in the area of multiview image-based modeling. The recent survey by Seitz et al. [3] gives an excellent review of the state-of-arts in this area. As summarized by [4], most of the existing algorithms follow a two-stage

approach: 1) conduct depth estimation based on local groups of input images; 2) fuse the estimated depth values into a global watertight 3D surface estimation. The depth estimation step is often based on image correlation [5]. The main differences between existing algorithms are in the second stage, the data fusion step, which can be divided into two categories. The first type of data fusion reconstructs the 3D surface by conducting volumetric data segmentation using global energy minimization approaches such as graph cut [6–8, 10, 17, 28], level-set [11–14, 23, 38, 39], or deformable models [5, 15, 16, 18]. Recently, people have proposed another types of data fusion algorithms that are based on local surface growing and filtering [2, 9, 10]. Without global optimization, these types of data fusion algorithms can be computationally more efficient [21, 22].

In this paper, we proposed an integrated depth fusion algorithm that integrates the graph-cut based global optimization with a mean-shift based explicit surface evolution. The energy functional to be minimized is derived from a novel volumetric saliency weighted normal vector field based on an anisotropic kernel density estimation. The experiments on the benchmark datasets show that our algorithm can obtain high quality reconstruction results that are comparable with the state-of-art methods, with considerable less computational time and complexity.

## 2 Algorithm

The entire algorithm consists of five main steps. Starting from an initial shape estimation such as the visual hull (Step 1), we will conduct depth estimation which will generate a set of 3D points representing the estimated depth (Step 2). A volumetric saliency weighted

---

Yongjian Xi  
xiyongjian@gmail.com

Ye Duan  
duanye@gmail.com

normal vector field is then constructed based on the 3D points (Step 3), which is then be used to extract a watertight 3D surface by graph cut algorithm (Step 4). Step 2 to Step 4 can be repeated several times. In practice, two to three iterations between Step 2 and Step 4 will be sufficient to obtain a good shape estimation, which will be further improved by the explicit surface evolution step (Step 5).

## 2.1 Saliency Weighted Normal Vector Field Construction

There are two main steps in the data fusion phase. First, a saliency weighted normal vector field is constructed based on the 3D points generated by the above depth estimation step. Next, a watertight 3D surface is extracted from the saliency weighted normal vector field by energy minimization. The saliency weighted normal vector field is constructed by the following three steps: 1) saliency field construction by anisotropic kernel density estimation; 2) normal estimation and consistent normal orientation propagation; 3) volumetric saliency weighted normal vector field construction.

We will detail the three steps of saliency weighted vector field construction in this section. The 3D surface extraction will be discussed in the later sections.

*Saliency field construction by anisotropic kernel density estimation:* We use the term saliency to represent the likelihood the unknown surface passes through a certain part of 3D space. In this paper, we propose to employ parzen-window based nonparametric density estimation method to compute the saliency of each point. Particularly, we propose to employ an anisotropic ellipsoidal kernel based density estimation method.

*Normal Estimation and Consistent Normal Orientation Propagation:* Given the 3D point clouds, we can estimate the normal vector at each point based on the Principle Component Analysis (PCA) algorithm [32]. Normal vectors estimated by the PCA algorithm however has an ambiguity of 180 degree so might not be consistently oriented. An orientation propagation is often needed to ensure the consistent orientation of the normal vectors. One way to do this is to first build a graph with each point as a node and the weights of edges between the adjacent points are defined as  $1 - \|n_1 \cdot n_2\|$ , where  $n_1$  and  $n_2$  are the normal vectors of the two adjacent points, and then compute the minimum spanning tree (MST) from the graph using algorithms such as the Kruskal's algorithm [33] which finds a subset of the edges that forms a tree that includes every vertex in the graph, where the total weight of all the edges in

the tree is minimized. At the termination of the algorithm, the normals are adjusted so the two neighbors in the tree have consistent normal orientation. The above MST based normal orientation propagation approach however is not robust against noises and outliers. Thus we propose to utilize external knowledge to guide the normal orientation propagation.

*Volumetric Saliency Weighted Normal Vector Field Construction:* Once we have estimated the saliency and normal vector at each point, we will proceed to construct a volumetric saliency weighted normal vector field, from which a watertight 3D surface can be extracted by energy minimization. A volumetric grid embedding all the 3D points is first constructed. The saliency and the normal vector of each point are then propagated to its adjacent grid nodes with weights calculated using the aforementioned parzen-window kernel function based on the anisotropic Mahalanobis distance between the grid node and the point.

## 2.2 3D Shape Estimation by Graph-Cut

Once the volumetric saliency weighted normal vector field is constructed, a watertight 3D surface  $S$  can be extracted by energy minimization. We use the following energy functional as suggested by [34]:

$$E = E_{data} + \varepsilon E_{reg} \quad (1)$$

$E_{data}$  is the data alignment term which is the inverse of the flux that enforces the surface alignment with the data orientation:

$$E_{data} = -flux(S) = - \int_S \langle N, v \rangle ds \quad (2)$$

where  $\langle \cdot, \cdot \rangle$  is (Euclidean) dot product and  $N$  is unit normal to surface elements  $ds$  consistent with a given orientation. If vectors  $v$  is interpreted as a local speed in a stream of water then the absolute value of flux equals the volume of water passing through the hypersurface in a unit of time. The sign of flux will be determined by the orientation of the surface.  $E_{reg}$  is the area-based regularization term that maintains the regularity of the extracted surface:

$$E_{reg} = \int_S ds \quad (3)$$

$\varepsilon$  is the coefficient of the regularization term  $E_{reg}$  that controls the strength of the smoothness in the energy minimization process and is related to the sampling density of the data. In our experiment, we set  $\varepsilon$  as 0.2.

As pointed out by [34], combining flux with area based regularization can overcome the shrinking effect

**Table 1** Running time and reconstruction accuracy.

Dataset	Running time (mins:secs)	# of input images	accuracy	completeness
Temple ring	33:50	47	99.5%	0.53
Temple sparse ring	29:15	16	96.8%	0.72
Dino ring	34:10	48	99.5%	0.46
Dino sparse ring	30:50	16	97.8%	0.42

of the area-based regularization and improve the reconstruction of elongated structures, narrow protrusions, and other fine details. Based on the divergence theorem for differentiable vector fields, the integral of flux of vector field over surface  $S$  equals to the integral of vector field's divergence  $div(v)$  in the interior of  $S$ :

$$\int_S \langle N, v \rangle ds = \int_V div(v_p) \cdot dp \quad (4)$$

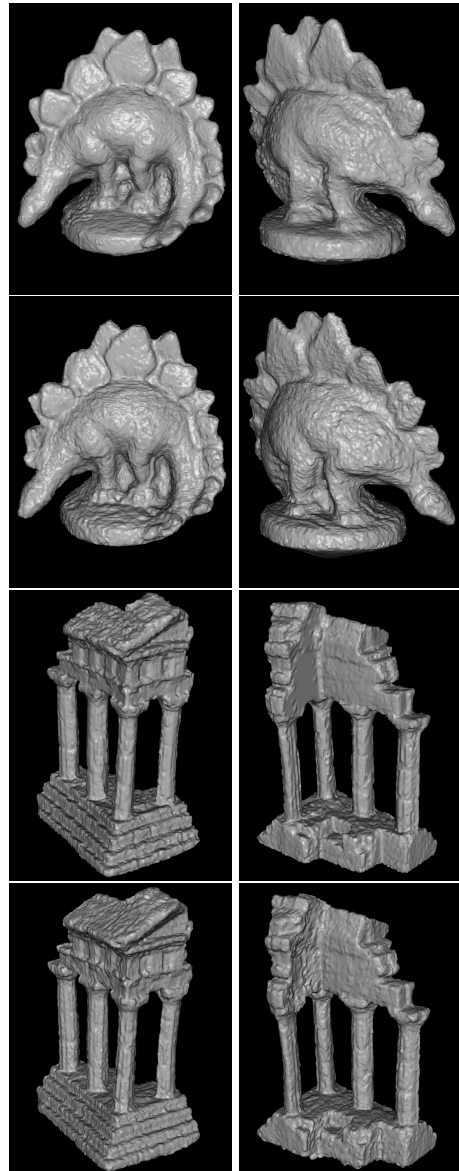
Where  $V$  is the region enclosed inside  $S$ . Thus  $E$  is now:

$$E = \varepsilon \int_S ds - \int_V div(v_p) \cdot dp \quad (5)$$

This equation can be solved efficiently using the graph cut algorithm [34]: neighboring nodes are connected via  $n$ -links representing area-based regularization cost. Nodes are also connected to the terminals via one  $t$ -link based on their divergence value: nodes have positive divergence are connected to the source terminal  $s$  with weight  $div(v_p)$ ; nodes have negative divergence are connected to the sink terminal  $t$  with weight  $-div(v_p)$ ; the node that has zero divergence is not connected to either terminal. The weight of the  $n$ -link is defined as the inverse of the edge length so that the weights of severed  $n$ -links approximate the surface area [35]. Consequently, a global minimum surface for the above equation can be found by computing a minimal  $s/t$ -cut in the constructed graph [34].

### 3 Benchmark Data Evaluation

We had applied our algorithm to the four benchmark datasets: temple ring, temple sparse ring, dino ring, and dino sparse ring from [19]. Table 1 shows the running time and the reconstruction accuracy and completeness obtained from the evaluation site [19]. The running time is based on a Pentium D Desktop PC with CPU 2.66GHz, 2GB RAM. The graph cut is conducted by using the generic max-flow library implemented by Boykov and Kolmogorov [37].



**Fig. 1** Reconstruction results of the four Benchmark datasets of [19]: dino sparse ring, dino ring, temple sparse ring, and temple ring.

### 4 Conclusion and Future Work

In this paper, we proposed: 1) a novel anisotropic kernel density based point saliency estimation algorithm; 2) a novel volumetric saliency weighted normal vector field; 3) a novel data fusion approach that integrates the global optimal graph cut algorithm with an efficient mean-shift based explicit surface evolution algorithm that can reconstruct the 3D surface with high-quality. The benchmark evaluation of our algorithm is among the top in the evaluation site, with considerable less computational time and complexity. Currently, our method utilizes the visual hull for initial estimation. In

the future we would like explore the use of bounding boxes and/or feature points as an initialization as were suggested by Furukawa et al. [36] and Quan et al. [24].

## References

1. Yang Wang, Xiaolei Huang, Chan-Su Lee, Song Zhang and Zhiguo Li, Dimitris Samaras, Dimitris Metaxas, Ahmed Elgammal, and Peisen Huang. High resolution acquisition, learning and transfer of dynamic 3D facial expressions. *Computer Graphics Forum*, 23(3):677-686, 2004
2. Michael Goesele, Noah Snavely, Brian Curless, Hugues Hoppe, Steven M. Seitz. Multi-View Stereo for Community Photo Collections, *Proceedings of ICCV 2007*, Rio de Janeiro, Brasil, October 14-20, 2007
3. Steve Seitz, Brian Curless, James Diebel, Daniel Scharstein, Richard Szeliski. A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms, *CVPR 2006*, vol. 1, pages 519-526
4. Neill Campbell, George Vogiatzis, Carlos Hernandez and Roberto Cipolla. Using multiple hypotheses to improve depth-maps for multi-view stereo. *ECCV 2008*, pages 766-779
5. C. Hernandez and F. Schmitt. Silhouette and stereo fusion for 3D object modeling. *CVIU*, 96(3):367-392, 2004
6. G. Vogiatzis, C. Hernandez, P.H.S. Torr, R. Cipolla. Multi-view stereo via volumetric graph-cuts and occlusion robust photo-consistency. *IEEE Trans. Pattern Anal. Mach. Intell.* 29(12) (2007)
7. M. Goesele, B. Curless, and S. Seitz. Multi-view stereo revisited. In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*. (2006)
8. A.Hornung, L. Kobbelt. Hierarchical volumetric multi-view stereo reconstruction of manifold surfaces based on dual graph embedding. In: *Proc. IEEE Conf. On Computer Vision and Pattern Recognition*. (2006)
9. Y. Furukawa, J. Pons. Accurate, dense, and robust multi-view stereopsis. In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*. (2007)
10. G. Vogiatzis, P. Torr, R. Cipolla. Multi-view stereo via volumetric graph-cuts. In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*. (2005)
11. H. Jin, S. Soatto, and A. Yezzi. Multi-view stereo reconstruction of dense shape and complex appearance. *IJCV*, 63(3):175-189, 2005
12. O. Faugeras and R. Keriven. Variational principles, surface evolution, PDE's, level set methods and the stereo problem. *IEEE Trans. on Image Processing*, 7(3):336-344, 1998
13. S. Soatto, A. Yezzi, and H. Jin. Tales of shape and radiance in multiview stereo. In *ICCV*, pp. 974-981, 2003
14. H. Jin, S. Soatto, and A. Yezzi. Multi-view stereo beyond lambert. In *CVPR*, vol. 1, pp. 171-178, 2003
15. Y. Duan, L. Yang, H. Qin, and D. Samaras. Shape reconstruction from 3D and 2D data using PDE-based deformable surfaces. In *ECCV*, vol. 3, pp. 238-251, 2004
16. C. Hernandez, F. Schmitt. Multi-stereo 3D object reconstruction, *Proceedings of 3D Data Processing Visualization and Transmission*, 2002. 159- 166
17. S. Sinha and M. Pollefeys. Multi-view reconstruction using photo-consistency and exact silhouette constraints: A maximum-flow formulation. In *ICCV*, pp. 349-356, 2005
18. Yasutaka Furukawa and Jean Ponce. Carved Visual Hulls for Image-Based Modeling, volume 1, pp. 564-577. *ECCV 2006*
19. The multi-view stereo evaluation web site at <http://vision.middlebury.edu/mview/>
20. M. Habbecke, L. Kobbelt. A surface-growing approach to multi-view stereo reconstruction. In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (2007)*
21. P. Merrell, A. Akbarzadeh, L. Wang, P. Mordohai, J-M. Frahm, R. Yang, D. Nister, M. Pollefeys. Real-time visibility-based fusion of depth maps. In: *Proc. 11th Intl. Conf. on Computer Vision*. (2007)
22. D. Bradley, T. Boubekeur, W. Heidrich. Accurate multi-view reconstruction using robust binocular stereo and surface meshing. In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*. (2008)
23. Maxime Lhuillier, Long Quan. A Quasi-Dense Approach to Surface Reconstruction from Uncalibrated Images. *IEEE Trans. Pattern Anal. Mach. Intell.* 27(3): 418-433 (2005)
24. Long Quan, Jingdong Wang, Ping Tan, Lu Yuan. Image-Based Modeling by Joint Segmentation. *International Journal of Computer Vision* 75(1): 135-150 (2007)
25. William E. Lorensen, Harvey E. Cline. Marching Cubes: A high resolution 3D surface construction algorithm. In: *Computer Graphics*, Vol. 21, Nr. 4, July 1987
26. A. Laurentini. The visual hull concept for silhouette-based image understanding. *IEEE Trans. Pattern Analysis and Machine Intelligence*. Pages: 150-162, 1994
27. Comaniciu, Dorin, Peter Meer. Mean Shift: A Robust Approach Toward Feature Space Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 24 (5): 603-619. 2002
28. V. Kolmogorov and R. Zabih. Multi-camera scene reconstruction via graph cuts. In *ECCV*, vol. III, pp. 82-96, 2002
29. Hong-Kai Zhao, S. Osher, R. Fedkiw. Fast surface reconstruction using the level set method, *Proceedings of IEEE Workshop on Variational and Level Set Methods in Computer Vision*, 2001. 194-201
30. Hong-Kai Zhao, S. Osher, B. Merriman, and M. Kang. Implicit and non-parametric shape reconstruction from unorganized points using variational level set method. *Computer Vision and Image Understanding*, 80(3):295-319, 2000
31. Vicent Caselles, Ron Kimmel, Guillermo Sapiro, Catalina Sbert. Three Dimensional Object Modeling via Minimal Surfaces. *ECCV (1) 1996*: 97-106
32. Hugues Hoppe, Tony DeRose, Tom Duchamp, John Alan McDonald, Werner Stuetzle: Surface reconstruction from unorganized points. *SIGGRAPH 1992*: 71-78
33. Joseph. B. Kruskal: On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem. In: *Proceedings of the American Mathematical Society*, Vol 7, No. 1 (Feb, 1956), pp. 48-50
34. Victor S. Lempitsky, Yuri Boykov: Global Optimization for Shape Fitting. *CVPR 2007*
35. Yuri Boykov, Vladimir Kolmogorov: Computing Geodesics and Minimal Surfaces via Graph Cuts. *ICCV 2003*: 26-33
36. Yasutaka Furukawa, Jean Ponce: Accurate, Dense, and Robust Multi-View Stereopsis. *CVPR 2007*
37. Yuri Boykov, Vladimir Kolmogorov: An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision, In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 9, pp. 1124-1137, Sept. 2008
38. Y. Liu, X. Cao, Q. Dai, and W. Xu. Continuous depth estimation for multi-view stereo. *CVPR 2009*
39. K. Kolev, T. Pock, and D. Cremers. Anisotropic minimal surfaces integrating photoconsistency and normal information for multiview stereo. *ECCV 2010*