Impact Factor-4.013 *e-ISSN: 2581-6667*
International Journal of Engineering and Creative Science, Vol. 3, No. 3, 2020
www.ijecs.net

Accurate Multiple View 3D Reconstruction Using Depth Merging

Ms. Roshani K. Dharme

Dept. of Computer Engineering G. H. Raisoni Polytechnic, Nagpur, India (M.S.)

Abstract – *This paper presents a quantitative comparison of several multi-view stereo reconstruction algorithms. In this paper, we propose a depth-map merging based hybrid method for large-scale scenes which takes both accuracy and efficiency is used into account. Different from typical multiview stereo methods, our approach is not only imposes the photo consistency constraint, but also associates explicitly the geometric coherence with multiple frames in a statistical way. An efficient stereo matching process will be used to generate depth-map at each image, followed by a depth-map refinement process to enforce consistency over neighboring views. Each depth-map is computed individually, which makes it appropriate for large scale scene reconstruction with high resolution images.*

Keywords- ***3D reconstruction, depth-map, multiple view stereo (MVS).***

I- INTRODUCTION

With fast developments of modern digital cameras, huge numbers of high declaration images could be easily captured nowadays. There is an urgent need to extract 3D structures from these images for many applications, such as architecture heritage protection, city-scale modeling, and so on.

Scene reconstruction from multiple images has always been an active field of research in computer vision. This classic problem finds many practical applications in the entertainment industry, in earth sciences and in cultural heritage digital archival for instance [29]. When high detail is needed, laser-based methods are usually applied successfully. However, these methods are rather complex to set for large-scale outdoor reconstructions, particularly when aerial acquisition is required.

These methods with image based ones, yielding considerable savings both in time and money. We believe that recent advances in multi-view stereo methods made this goal closer than ever.

II - RELATED WORK

Large scale reconstruction systems typically generate partial reconstructions which are then merged. Conflicts and errors in these partial reconstructions are identified and resolved during the merging process. Multiple-view reconstruction methods based only on images have also been thoroughly investigated [1], but many of them are limited to single objects and cannot be applied to large scale scenes due to computation and memory requirements.

Turk and Levoy [2] proposed a method for registering and merging two triangular meshes. They remove any overlapping parts of the meshes, connect the mesh boundaries and then update the positions of the vertices. Soucy and Laurendeau [3] introduced a similar algorithm which first updates the positions of the vertices and then connects them to form the triangular mesh. A different approach was presented by Curless and Levoy [4] who employ a volumetric representation of the space and compute a cumulative weighted distance function from the depth estimates. This signed distance function is an implicit representation of the surface. A volumetric approach that explicitly takes into account boundaries and holes was published by Hilton et al. [5]. The voxel based methods compute a cost

function on a 3D volume which is a bounding box of the object. Seitz et al. [6] propose a voxel coloring framework that traverses a discrete 3D space in a generalized depth-order to identify voxels that have a unique color, constant across all possible interpretations of the scene. Vogiatzis et al. [7] use graph-cut optimization to compute the minimal surface that encloses the largest possible volume, where surface area is just a surface integral in this photo- consistency field. Goesele et al. [8] use Normalized Cross Correlation (NCC) based pixel window matching techniques to produce depthmaps then merge them with volumetric integration. Strecha et al. [9] jointly model depth and visibility as a hidden Markov Random Field, and use EM-algorithm to optimize the model parameters. Merrell et al. [10] first use a computationally cheap stereo algorithm to generate potentially noisy, overlapping depth-maps, and then fuse these depth-maps to obtain an integrated surface based on visibility relations between points. Zach et al. [11] present a method to globally optimize an energy functional consisting of a total variation regularization force and an $L1$ data fidelity term.

Bradley et al. [12] propose a method which uses robust binocular stereo with scaled matching windows, followed by adaptive point-based filtering of the merged point clouds. Campbell et al. [13] store multiple depth hypotheses for each pixel and use a spatial consistency constraint to extract the true depth in the discrete Markov Random Field framework. Liu et al. [14] produce high quality MVS reconstruction results using continuous depth-maps generated by variation optical flow. But this method requires the visual hull as an initialization. Li et al. [15] generate depth-maps using DAISY [16] feature, and use two stages of bundle adjustment to optimize the positions and normal of 3D points. Tola et al. [17] also use DAISY feature to generate depth-maps, and then merge them by consistency checking at neighboring views. This method is similar to our method, but ours uses patch based stereo instead of merely matching DAISY features along epipolar lines, which can produce more accurate depth-maps without diminishing the computational efficiency.

III - OVERVIEW OF METHOD

To develop an efficient stereo matching process is used to generate depth-map at each image with acceptable errors, followed by a depth-map refinement process to enforce consistency over neighboring views.

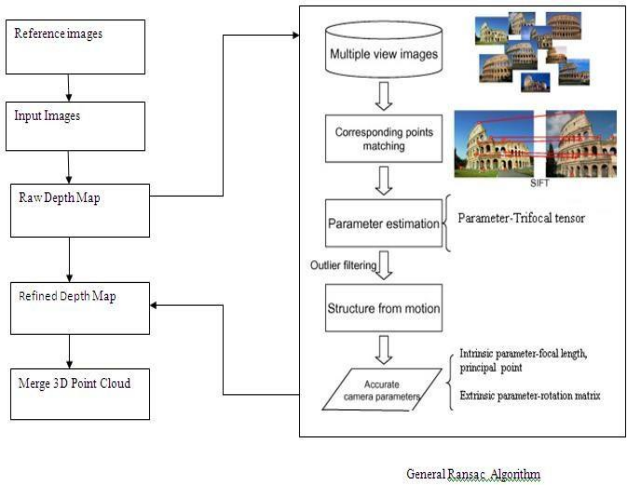

Fig. 1 Framework of the Method

In Fig.1 for each image in the input image set, a reference image to form a stereo pair for depth map computation. Since these raw depth-maps generated by stereo vision may contain noises and errors. It refine each of them by consistency checking using its neighboring depth maps. Finally all the refined depth-maps are merged together to get a final reconstruction.

IV - SCENE ANALYSIS

The geometry of an object or scene can be represented in numerous ways; the vast majority of multi-view algorithms use voxels, level-sets, polygon meshes, or depth maps.

Many techniques represent geometry on a regularly sampled 3D grid (volume), either as a discrete occupancy function or as a function encoding distance to the closest surface. 3D grids are popular for their simplicity, uniformity, and ability to approximate any surface.

Polygon meshes represent a surface as a set of connected planar facets. This multi-depth map representation avoids resembling the geometry on a 3D domain, and the 2D representation is convenient particularly for smaller datasets. An alternative is to define the depth maps relative to scene surfaces to form a relief surface.

- **Stereo Pair Selection**

For each image in the image set, we need to select a reference image for it for stereo computation. The selection of stereo image pair is important not only for the accuracy of the stereo matching but also for the final

MVS result. Stereo pair selection is a relatively easy task for street-side view cameras on the vehicle [18]–[21] or cameras in a controlled environment like the Middlebury benchmark data [1], but needs to be carefully designed for unordered images. A *good* candidate reference image should have a similar viewing direction as the target image, and have a suitable baseline neither too short to degenerate the reconstruction accuracy nor too long to have less common coverage of the scene.

V - MULTI-VIEW STEREO

Since the review of Seitz et al. [1] and the associated Middlebury evaluation, a lot of research has been focusing on multi-view reconstruction of small objects with tightly controlled imaging conditions. This has led to the development of many algorithms whose results are beginning to challenge the precision of laser-based reconstructions. However, as we will see, most of these algorithms are not directly suited to large-scale scenes. A number of multi-view stereo algorithms have been proposed that exploit the visual hull [24]. They rely on it either as an initial guess for further optimization [27], as a soft constraint or even as a hard constraint [27] to be fulfilled by the reconstructed shape. While the unavailability of the visual hull discards many of the top-performing multi-view stereo algorithms of Middlebury challenge for our purpose, the requirement for the ability to handle large-scale scenes discards most of the others, in particular volumetric methods, i.e. methods based on a regular decomposition of the domain into elementary cells, typically voxels. Finally, cluttered scenes disqualify variational methods [26] that get stuck into local minima, unless they provide a way of estimating a close and reliable initial guess that takes visibility into account.

VI - IMAGE PREDICTION USING DEPTH MAPS

Depth Map Estimation

Exact disparity compensation can be performed only if scene depth is accurately known for each image pixel. Many different, and often computationally demanding, algorithms have been proposed to estimate depth from stereoscopic as well as multi-view images. The derivation of correct depth at pixel resolution, however, proves to be an elusive problem. Fortunately, with regard to disparity-compensated image prediction, determining true scene depth is not imperative. Rather, the depth value yielding the best prediction result must be found. To estimate dense depth maps, the multi-view images are divided into blocks. As depicted in Fig. 2, each image block is compared to the reference images used later for prediction.

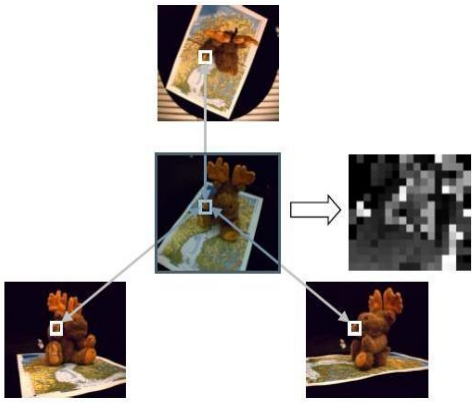

Fig. 2 Block-based depth map estimation

- **Depth-Map Computation**

For each eligible stereo pair, we follow the idea in [23] to compute the depth-map. The core idea is that, for each pixel in the input image we try to find a good support plane that has minimal aggregated matching cost with the reference image, as shown in Fig.3. The support plane *f* is essentially a local tangent plane of the scene surface, which is represented by a 3D point *Xi* and its normal *ni* in the related camera's coordinate system, as shown in Fig. 4.

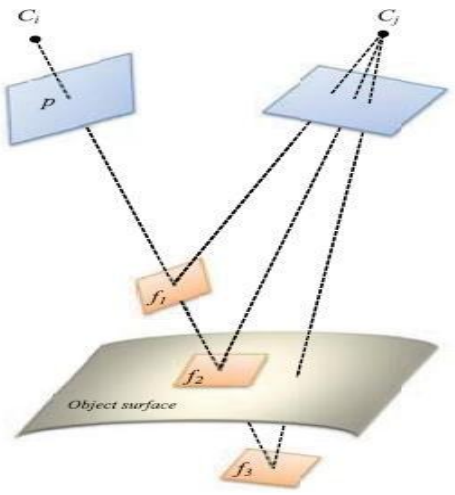

Fig 3**.** For each pixel *p* in the input image, we estimate its Corresponding

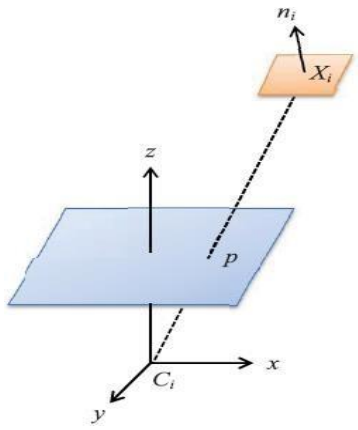

Fig. 4.Support plane is represented by a 3D point *Xi* and its normal *ni* in camera *Ci* 's coordinate

- **Depth-Map Refinement**

Since the raw depth-maps may not completely agree with each other on common areas due to depth errors, a refinement process is carried out to enforce consistency over neighboring views. For each point *p* in image *Ii* , we back project it to 3D using its depth *λ* and the camera parameters, as:

$$X = \lambda RT\, i\, K{-}1\, i\, p + Ci$$

After the above refinement process, most errors could be removed, which results in a relatively clean depth-map at each view.

- **Depth-Map Merging**

After refinement, all the depth-maps could be merged to represent the scene. However, merging depth-maps directly may contain lots of redundancies, because different depthmaps may have common coverage of the scene, especially for neighboring images. In order to remove these redundancies, the depth-maps are further reduced by neighboring depthmap test. As illustrated by Fig. 5, for each pixel in camera.

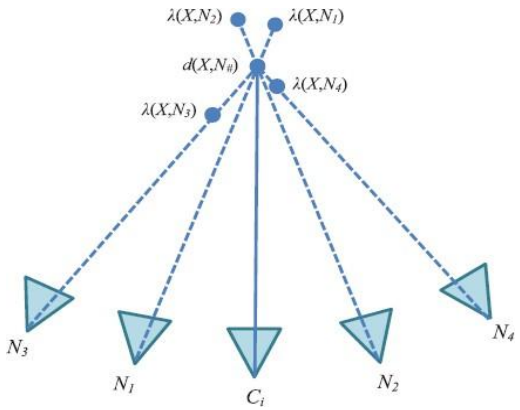

Fig. 5. Illustration of redundancy removing by depth-map test

Finally, all depth-maps are back projected into 3D, and merged into a single point cloud. The final point cloud is usually quite dense especially when using high resolution images. If we want to make it sparse, we can simply just back project points at sparse locations in the depth-maps. For example, using only points at image locations *(2n, 2n)* in the depth-map will approximately reduce the size of the point cloud to a quarter of the size that use all points. This gives us a way to control the point cloud size according to memory and storage limitations.

VII - RANDOM SAMPLE CONSENSUS (RANSAC)

RANSAC algorithm proposed by Fischler and Bolles [28] is a general parameter estimation approach designed to cope with a large proportion of outliers in the input data. Unlike many of the common robust opinion techniques such as M- estimators and least-median squares that have been adopted by the computer vision community from the information literature, RANSAC was developed from within the computer vision community. RANSAC is a similar to technique that generate candidate solutions by using the minimum number observations (data points) required to estimate the underlying model parameters. As pointed out by Fischler and Bolles [28], unlike conventional sampling techniques that use as much of the data as possible to obtain an initial solution and then proceed to reduce outliers, RANSAC uses the smallest set possible and proceeds to enlarge this set with consistent data points [28].

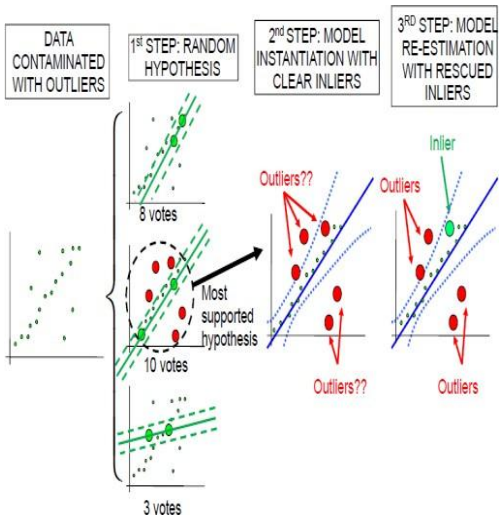

Fig 6: RANSAC steps for the simple 2D line estimation

In order to highlight the requirements and benefits of our method, the RANSAC algorithm [26] is first briefly exposed in this introduction for the simple case of 2D line estimation from a set of points contaminated with spurious data (see Fig.6

The basic algorithm is summarize as follows: Algorithm 1 RANSAC

1: Select randomly the minimum number of points required to determine the model parameters.

2: Solve for the parameters of the model.

3: Determine how many points from the set of all points fit with a predefined tolerance e

4: If the fraction of the number of inliers over the total number points in the set exceeds a predefined threshold t re-estimate the model parameters using all the identified inliers and terminate.

5: Otherwise, repeat steps 1 through 4 (maximum of N times).

VIII - CONCLUSIONS

The propose depth-map merging based method and RANSAC algorithm for large scale scenes which takes both accuracy and efficiency into account**.**. A novel RANSAC algorithm is presented in this paper which, for the first time and differently from standard purely data-driven RANSAC, incorporates a priori probabilistic information into the hypothesis generation stage. As a consequence of using this prior information, the sample size for the hypothesis generation loop can be reduced to the minimum size of 1 point data. It could be easily parallelized at image level, i.e., each depth-map is computed individually, which makes it suitable for large-scale scene reconstruction with high resolution images.

REFERENCES

[1] *S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski,." A comparison and evaluation of multi-view stereo reconstruction algorithms," In CVPR, pages 519–528, 2006.*

[2] *G. Turk and M. Levoy. "Zippered polygon meshes from range images,"In SIGGRAPH, pages 311–318, 1994.*

[3] *M. Soucy and D. Laurendeau,"A general surface approach to the integration of a set of range views,"PAMI, 17(4):344–358, 1995.*

[4] *B. Curless and M. Levoy," A volumetric method for building complex models from range images," SIGGRAPH, 30:303– 312, 1996.*

[5] *A. Hilton, A. Stoddart, J. Illingworth, and T. Windeatt. "Reliable surface reconstruction from multiple range images," In CVPR, pages 117–126, 1996.*

[6] *S. M. Seitz and C. R. Dyer, "Photorealistic scene reconstruction by voxel coloring," Int. J. Comput. Vis., vol. 35, no. 2, pp. 151–173, Nov. 1999.*

[7] *G. Vogiatzis, C. Hernandez, P. H. Torr, and R. Cipolla, "Multiview stereo via volumetric graph-cuts and occlusion robust photo-consistency," IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 12, pp. 2241–2246, Dec. 2007.*

[8] *M. Goesele, B. Curless, and S. M. Seitz, "Multi-view stereo revisited," in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., Oct. 2006, pp. 2402–2409.*

[9] *C. Strecha, R. Fransens, and L. V. Gool, "Combined depth and outlier estimation in multi-view stereo," in Proc. IEEE Comput. Soc. Conf.Comput. Vis. Pattern Recognit., Oct. 2006, pp. 2394–2401.*

[10] *P. Merrell, A. Akbarzadeh, L. Wang, P. Mordohai, and J.-M. Frahm,"Real-time visibility-based fusion of depth maps," in Proc. IEEE Int.Conf. Comput. Vis., Oct. 2007, pp. 1–8.*

[11] *C. Zach, T. Pock, and H. Bischof, "A globally optimal algorithm for robust tv-l1 range image integration," in Proc. IEEE Int. Conf. Comput.Vis., Oct. 2007, pp. 1–8.*

[12] *D. Bradley, T. Boubekeur, and W. Heidrich, "Accurate multi-view reconstruction using robust binocular stereo and surface meshing," in Proc.IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., Jun. 2008,pp. 1–8.*

[13] *N. D. Campbell, G. Vogiatzis, C. Hernandez, and R. Cipolla, "Using multiple hypotheses to improve depth-maps for multi-view stereo," in Proc. Eur. Conf. Comput. Vis., Oct. 2008, pp. 766–779.*

[14] *Y. Liu, X. Cao, Q. Dai, and W. Xu, "Continuous depth estimation for multi-view stereo," in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., Jun. 2009, pp. 2121–2128.*

[15] *J. Li, E. Li, Y. Chen, L. Xu, and Y. Zhang, "Bundled depth-map merging for multi-view stereo," in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., Aug. 2010, pp. 2769–2776.*

[16] *E. Tola, V. Lepetit, and P. Fua, "Daisy: An efficient dense descriptor applied to wide-baseline stereo," IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 5, pp. 815–830, May 2010.*

[17] *E. Tola, C. Strecha, and P. Fua, "Efficient large- scale multi-view stereo for ultra high-resolution image sets," Mach. Vis. Appl., vol. 23, no. 5, pp. 903–920, 2012.*

[18] *D. Gallup, J.-M. Frahm, P. Mordohai, Q. Yang, and M. Pollefeys, "Realtime plane-sweeping stereo with multiple sweeping directions," in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., Jun. 2007, pp. 1–8.*

[19] *M. Pollefeys, D. Nister, J.-M. Frahm, A. Akbarzadeh, P. Mordohai, B. Clipp, C. Engels, D. Gallup, S. J. Kim, P. Merrell, C. Salmi, S. N. Sinha, B. Talton, L. Wang, Q. Yang, H. Stewenius, R. Yang, G. Welch, and H. Towles, "Detailed real-time urban 3D reconstruction from video," Int. J. Comput. Vis., vol. 72, no. 2, pp. 143–167, 2008.*

[20] *G. Zhang, J. Jia, T.-T. Wong, and H. Bao, "Consistent depth maps recovery from a video sequence," IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 6, pp. 974–988, Jun. 2009.*

[21] *D. Gallup, J.-M. Frahm, and M. Pollefeys, "Piecewise planar and nonplanar stereo for urban scene reconstruction," in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., Jun. 2010, pp. 1418–1425.*

[22] *Marcus Magnor, Peter Eisert," Multi-View Image Coding with Depth Maps and 3-D Geometry for Prediction," SPIE Conference Proceedings: Visual Communications and Image Processings (VCIP-2001) San Jose, CA, pp. 273-271, January 2001*

[23] *M. Bleyer, C. Rhemann, and C. Rother, "Patchmatch stereo—stereo matching with slanted support windows," in Proc. Brit. Mach. Vis. Conf., Aug.–Sep. 2011, pp. 14.1–14.11.*

[24] *A. Laurentini, "The visual hull concept for silhouette-based image understanding," IEEE Transactions on Pattern Analysis and Machine Intelligence, 16(2):150–162, 1994.*

[25] *Y. Furukawa and J. Ponce,"Carved visual hulls for imagebased modeling," In European Conference on Computer Vision, 2006.*

[26] *C. Hern´andez and F. Schmitt," Silhouette and stereo fusion for 3D object modeling. Computer Vision and Image Understanding," 96(3):367–392, 2004.*

[27] *P. Labatut, J.-P. Pons, and R. Keriven,"Robust and efficient surface reconstruction from range data," submitted to Computer Graphics Forum, 2009.*

[28] *M.A. Fischler and R.C. Bolle, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," Communications of the ACM, 24(6):381–395, 1981.*

[29] *Vu Hoang Hiep, Renaud Keriven,Patrick Labatut," Towards high-resolution large-scale multi-view stereo," Universit´e Paris-Est, LIGM/ENPC/CSTB.*