Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2018) 26: 755 – 767

doi:10.3906/elk-1704-144

Volumetric 3D reconstruction of real objects using voxel mapping approach in a multiple-camera environment

Tushar JADHAV[1,*], **Kulbir SINGH**[2], **Aditya ABHYANKAR**[3]
[1]Department of Electronics & Telecommunication, Vishwakarma Institute of Information Technology, Pune, India
[2]Department of Electronics & Communication Engineering, Thapar University, Patiala, India
[3]Department of Technology, Savitribai Phule Pune University, Pune, India

Received: 12.04.2017 • **Accepted/Published Online:** 13.12.2017 • **Final Version:** 30.03.2018

Abstract: Extracting 3D information from 2D images is an inverse estimation problem and a challenging task in itself. The aim of 2D to 3D reconstruction is to generate either a volume or a surface representing the object from multiple views. This paper presents a simple and accurate multiple-view volumetric 3D reconstruction method using an integrated approach based on homography estimation and voxel mapping. The homography-based approaches give accurate estimates but do not provide system dynamics. The voxel-based volumetric reconstruction methods provide system dynamics that are essential for system modeling. However, they face challenges while modeling the concavities. This paper presents a proposed 3D reconstruction method that combines homography estimation and the voxel mapping approach for improving the accuracy of 3D reconstruction. Experimental results show that the method efficiently reconstructs objects of known and unknown shape, fragile objects, and complex scenes with multiple objects. The use of homography along with voxel mapping in a multiple-camera environment brings out more details of the object for improving the quality of reconstruction.

Key words: Voxel mapping, volumetric reconstruction, homography, camera calibration, multiple-camera environment

1. Introduction

Three-dimensional reconstruction is an integral part of various applications in almost all disciplines. Some important applications of this field include virtual reality, robot navigation, augmented reality tasks, games, animation, motion pictures, industrial measurements, surface analysis, volumetric analysis, and forensic applications [1]. Three-dimensional reconstruction is an inverse estimation problem and 3D information can be recovered from 2D images if the camera projection matrices are known. Several attempts have been made by researchers to reconstruct the 3D shape of an object from multiple images. There is always a tradeoff between computation speed, computation complexity, accuracy, and feasibility in the implementation of these methods. Consequently, bringing more accuracy and quality to 3D reconstruction is still a challenging and open research problem. The homography-based approaches give accurate estimates but do not provide system dynamics. Hence, these approaches are used in applications such as tennis referral systems, where system dynamics are not important [2–4]. On the other hand, the voxel-based volumetric reconstruction methods provide system dynamics that are essential for system modeling. These methods provide the true essence of the volume of an object. However, these approaches face challenges while modeling the concavities. This paper presents the

*Correspondence: tushar.jadhav@viit.ac.in

proposed multiple-view volumetric 3D reconstruction method, which combines homography estimation and the voxel mapping approach for improving the accuracy of 3D reconstruction.

The rest of the paper is organized as follows. Section 2 presents the related works in the field of 3D reconstruction. Section 3 throws light on the methodology and details of the proposed 3D reconstruction method. Experimental results of the proposed method are discussed in Section 4. Section 5 presents the conclusion.

2. Related works

The literature reveals that many researchers have contributed to the development of 3D reconstruction over the last two decades. Researchers have used voxels, polygon mesh, or level sets for representing objects [5]. Algorithms based on the multi-view stereo (MVS) match feature points in multiple images of the object and compute depth maps. Later the depth maps are fused into a single depth map to extract the object surface [6]. These approaches are less efficient computationally and are complex for implementation. The surface reconstruction method designed for multiscale scenes uses very dense matching of feature points in multiple images to reconstruct the object surface [7]. The approach developed by Jang et al. [8] uses multiple color images along with the depth image to extract the 3D surface and generate the 3D model. Although the method handles the self-occlusion problem, further improvement is needed to make the method robust. The 3D surface reconstruction method proposed by Shen [9] makes use of a patch-based stereo matching approach. The method is able to handle large-scale scenes and produce dense point clouds. However, it is less efficient computationally.

The approaches using voxel-based volumetric representation obtain the visual hull of an object from its silhouettes [10]. Although the method addresses the issue of illumination constraints, the algorithm is computationally complex. The method developed by Srinivasan et al. [11] uses geometric intersection to obtain the visual hull of an object and extract the surface by employing a contour model. However, the algorithm faces challenges of constrained resolution while constructing these contours. A survey of volumetric 3D reconstruction methods was presented in [6]. The approach presented in [12] uses multiresolution mapping for volumetric reconstruction. The method using the silhouette probability maps deals with the errors in camera calibration or the errors in the silhouette generation step robustly [13]. Although the algorithm handles objects with less texture, further improvement in the accuracy of reconstruction is needed. The approach developed by Chen et al. [14] uses the depth map fusion approach to obtain a single volumetric representation for an object and extract the surface. The algorithm addresses the issue of computation time/memory tradeoff. Development of the MVS methods of 3D reconstruction has witnessed tremendous growth in recent years [15]. The approach using the deep learning algorithm uses 3D CAD models and reconstructs objects from a single depth map [16]. Another learning-based approach for 3D reconstruction uses the relation between 2D observation and object shape [17]. However, the learning step makes these approaches more complex computationally. The surface representations obtained by these methods are often incomplete and sensitive to numerical instabilities. Moreover, they face challenges while reconstructing complex objects. On the other hand, the volumetric methods are able to reconstruct 3D models of objects with a complex shape [18]. Zhao and Xiao [19] implemented the voxel coloring algorithm in the HLS color space instead of the RGB color space in order to overcome the limitation of the ordinary voxel coloring algorithm. However, the conversion of multiple images into HLS color space increases the computation time. An automatic reconstruction of the volume is possible with the parameterization of a voxel space and calculation of the 3D polyhedron [20]. However, this makes the algorithm complex. Homography-based methods are widely used in tracking applications. Though accurate, these methods do not give the

system dynamics needed for modeling [3]. The proposed method integrates homography estimation and the voxel mapping approach for improving the accuracy of 3D reconstruction.

3. Methodology

The proposed 3D reconstruction method uses multiple cameras to acquire multiple images of an object. Figure 1 presents the schematic diagram of the reconstruction pipeline of the proposed method. After acquiring multiple images of the objects, images of the calibration grid are obtained for camera calibration. The camera parameters are used to compute camera matrices and to determine the initial limits of the volume of the object. Subsequently, the reconstruction algorithm performs voxel mapping based on the silhouette and homography estimates. The last step in the proposed method is to extract the surface of the object. Details of the reconstruction pipeline are presented in the following subsections.

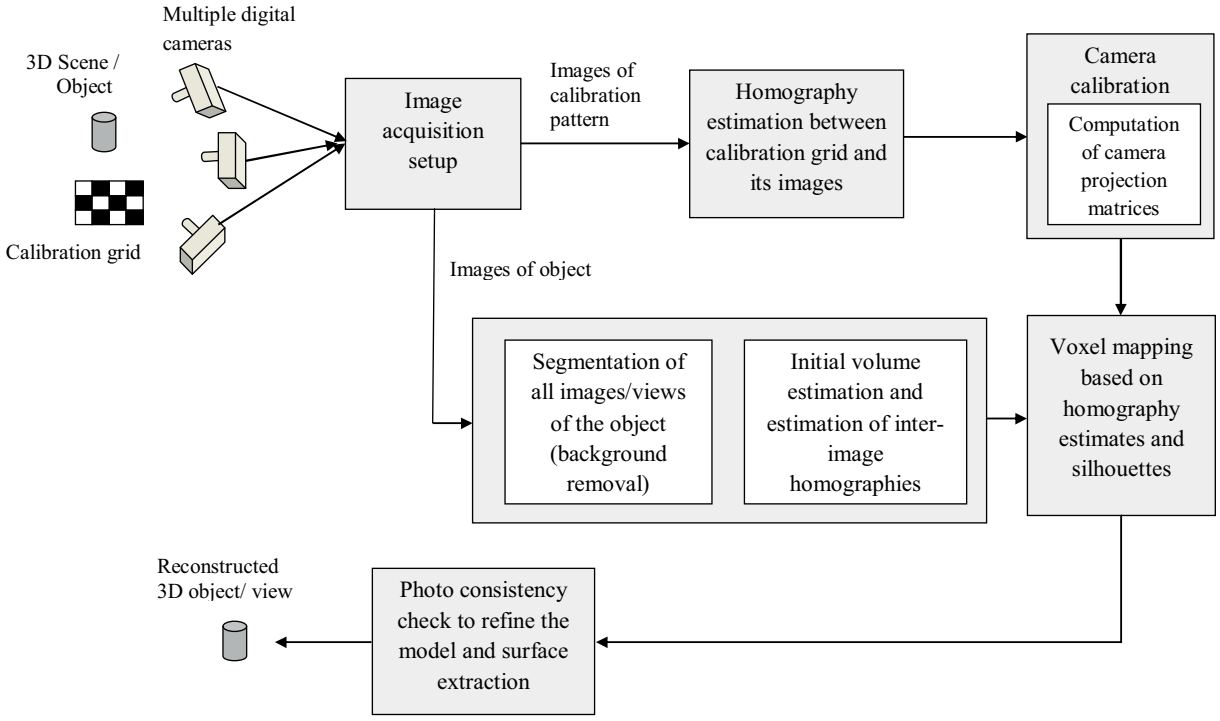

Figure 1. Schematic diagram of the reconstruction pipeline of the proposed method.

3.1. Image acquisition and camera calibration

The proposed 3D reconstruction method uses 4 Nikon-S3600 digital cameras. The cameras are placed at different elevations around the turntable for acquiring multiple images of the calibration pattern and of the objects to be reconstructed. All the images are identical in size (1600×1200) and resolution (300 dpi). The accuracy of camera calibration governs the accuracy of the reconstruction method. It is the process of computing the internal parameters, such as focal length, principal point, skew and external parameters, translation vector, and rotation matrix of the camera. These parameters are used further to compute the camera matrices and pose of the cameras. While imaging, the camera maps the object point, $P(X,Y,Z)$, to the corresponding 2D image point, $P_c(uv)$ [21]. Mathematically, this transformation is expressed as follows:

$$P \mapsto P_c. \tag{1}$$

Mapping Eq. (1) can be expressed in homogeneous coordinates as:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = [K_{3\times3}]. \begin{bmatrix} R & -Rt \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \tag{2}$$

where $K_{3\times3}$ is the intrinsic parameter matrix of the camera:

$$K_{3\times3} = \begin{bmatrix} m_u f & s & m_u t_u \\ 0 & m_v f & m_v t_v \\ 0 & 0 & 1 \end{bmatrix}, \tag{3}$$

where f represents the focal length, $m_u t_u$ and $m_v t_v$ represent the principal point coordinates, s represents the skew parameter, and m_u and m_v represent the pixel resolution along the axes of the image plane of the camera. Translation vector $t_{3\times1}$ and rotation matrix $R_{3\times3}$ are the extrinsic parameters of the camera. Thus, the final mapping equation is:

$$P_c = C_{3\times4} P, \tag{4}$$

where the camera matrix is:

$$C_{3\times4} = [K_{3\times3}]\,[R_{3\times3}]\,[I_{3\times3}, -t_{3\times1}]\,. \tag{5}$$

The camera matrix consists of twelve elements, of which eleven elements are unknown variables and the last element is unity. The problem of finding the eleven unknown variables of the camera matrix requires the set of linear equations giving correspondences between the 3D points and respective image points. The camera calibration toolbox developed by Bouguet is used for this purpose [22,23]. A calibration grid sized 270 mm $\times$ 210 mm with each square sized 25 mm $\times$ 25 mm is used as the calibration object. Figure 2 shows sample images of the calibration grid acquired using one of the cameras in a multiple-camera setup. The images are acquired by rotating a calibration grid using the turntable while maintaining the same pose of the camera. After selecting the tie points, the calibration toolbox detects the corner points of the calibration grid and estimates the camera parameters and the camera matrix. All the cameras are calibrated using same procedure. Distortion parameters of the cameras are neglected due to their small value.

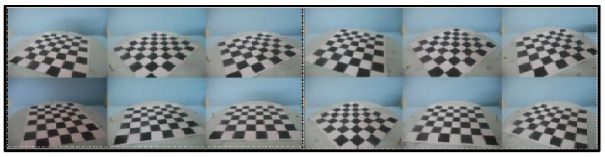

Figure 2. Images of calibration grid acquired using a camera in a multiple-camera setup.

3.2. The 3D reconstruction

The proposed method combines homography estimation and the voxel mapping approach. The method uses object silhouettes along with homography estimates to decide the occupancy of the voxel. In the beginning, the method estimates the limits of the initial volume derived from the poses of multiple cameras. The proposed method uses voxel representation for modeling the object. Hence, an initial volume is discretized to form

the 3D voxel grid. The images acquired in the image acquisition step are segmented using the background subtraction method to obtain object silhouettes. The background subtraction approach used in this method is also suitable for complex backgrounds. However, a uniform and plain background is maintained during the experimentation in order to make the segmentation process simpler. The object silhouettes are used to estimate the silhouette probabilities in the multiple-camera environment. The succeeding step computes the interimage homographies. The interimage homographies are estimated using direct linear transformation and singular value decomposition to obtain relevant silhouettes for each view position. Each voxel is projected using the camera matrices on corresponding images. The homographies are used to warp these images on the reference image and the error between the 2D projections of warped points and the 2D projection in reference view is computed. Lower error signifies better correlation between the views. The error estimate is further used along with the silhouette probabilities to decide the occupancy of the voxel. In the voxel mapping step, the voxels are mapped either to the object or to the background based on their occupancy weights. The occupancy weights of the voxels are used further as isovalues in the surface extraction process. The model is refined later by applying the color consistency constraint. This process is carried out for all voxels in the 3D voxel grid. The last step in the reconstruction pipeline is to extract the object surface from the voxel volume. The proposed method uses a surface triangulation approach to reconstruct the surface from the volume data efficiently. The reason for the use of the surface triangulation approach is to overcome the limitations of the voxel and particle representations of the surface. The voxel representations face problems of the visibility of voxel cubes after zooming, whereas the particle representations face problems of the holes in the reconstructed surface [24]. Here, each grid cell is represented in terms of its vertices and corresponding scalar values. Based on the isovalue, the surface extraction process creates further planar facets representing the isosurfaces passing through the grid cell. Subsequently, triangular facets corresponding to the isosurface with unique codes are generated. Each grid cell with eight vertices can have a maximum of 256 configurations of triangular facets. The grouping of symmetric configurations reduces the number of configurations to 15. The vertices having scalar values below or above the isovalue are identified to assign an 8-bit index to the voxel. This index is mapped to a unique 12-bit code using the edge look-up table, which gives the information about the edges that the corresponding isosurface intersects. The vertices of the isosurface that lie on the edges are obtained using the linear interpolation. Let us consider one of the edges of the voxel cube that joins vertex A and vertex B with corresponding scalar values 'V_1' and 'V_2'. The isosurface with the isovalue 'V' intersects the edge at point 'T', which is determined as follows:

$$T = \left\{ \frac{A\,(V - V_2)}{V_2 - V_1} + \frac{B\,(V - V_1)}{V_2 - V_1} \right\} \tag{6}$$

This leads to the computation of triangle vertices and their normals to further extract the object surface. The surface mesh model of a toy object, a duck, showing a magnified view of the mesh corresponding to the concavity is presented in Section 4.

4. Experimental results and discussion

The experimentation is performed on an Intel Core i5-2450M CPU @ 2.50 GHz using the Windows 7 operating system. The proposed method is tested using objects of known and unknown shapes. Objects with conventional shapes, such as spheres, prisms, and cubes, are used as objects with known shape and geometrical parameters, while toy objects such as an orange fish, duck, rooster, rabbit, and wooden pot are used as objects having shapes different than the conventional geometrical shapes. The shape of the wooden pot is less complex than the shape of the other objects. The performance of the method is evaluated by comparing the volume obtained

through 3D reconstruction for sphere-, prism-, and cube-shaped objects with their actual volume obtained using geometrical measurements and volume measured using the water displacement method. Figure 3 shows 9 sample views out of 72 views of sphere-, prism-, and cube-shaped objects [25] acquired using the multiple-camera setup. Figure 4 shows 9 sample views of the orange fish, duck, rooster, rabbit, and wooden pot acquired using the multiple-camera setup.

Figure 3. Nine sample views out of 72 views of the objects with known shapes, sphere, prism, and cube [25], acquired using a multiple-camera setup.

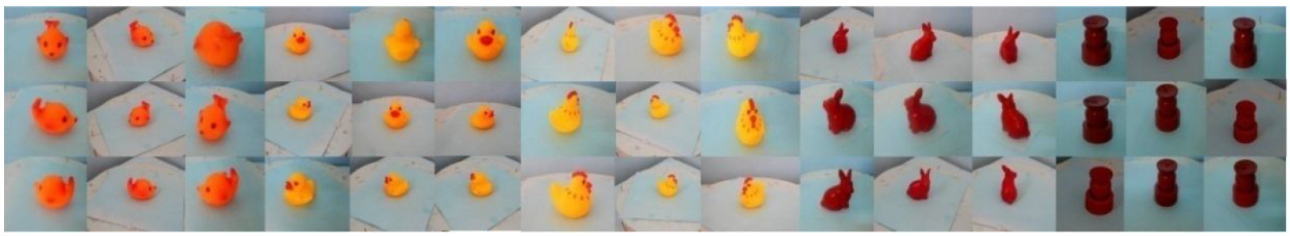

Figure 4. Nine sample views out of 72 views of the objects, orange fish, duck, rooster, rabbit, and a wooden pot, acquired using a multiple camera setup.

The binary silhouettes of the objects with known shapes and the objects with unknown shapes are shown in Figure 5 and Figure 6, respectively. The silhouettes are further used along with the homography estimates in the voxel mapping process and to find the voxel occupancies. Results show that the quality of the silhouettes is good. Hence, a voting approach is not needed while obtaining the silhouettes. Figure 7 shows the surface mesh model of the 3D reconstruction of duck. Results show that the method models the concavities efficiently. Figures 8 and 9 show the results of the 3D reconstruction of objects with known shapes and objects with unknown shapes using the proposed method respectively. Results show that the proposed method reconstructs the objects with known and unknown shapes efficiently. The method is also tested for reconstruction of a fragile object and of a scene with two objects. The sample images, silhouettes, and reconstruction results in both the cases are shown in Figure 10. From Figures 8–10, it is evident that the method reconstructs the objects with complicated shapes with concavities efficiently. Thus, the use of homography estimates and silhouettes for computation of the voxel occupancies in a multiple-camera environment helps the method to improve the reconstruction quality. Details such as the beak and wings of the duck, wings and tail of the orange fish, wattles and comb points of the rooster, and overall shape of the rabbit are reconstructed with good quality. The method reconstructs the objects having few feature details, such as the wooden pot, efficiently. Figure 11 shows several views of the 3D reconstructed objects after texture mapping.

Table 1 presents the comparison between the volume of the objects with known shapes obtained using the geometrical parameters, their volume measured using the water displacement method, and the volume estimated using the proposed method. Table 2 presents the comparison between the real-world volume and the estimated volume of objects with unknown shapes. The estimated volume is obtained by converting the voxel volume into a real-world unit (cm^3). The voxel resolution and the camera parameters are used for converting the volume

Figure 5. Sample silhouettes of the objects with known shapes: sphere, prism, and cube.

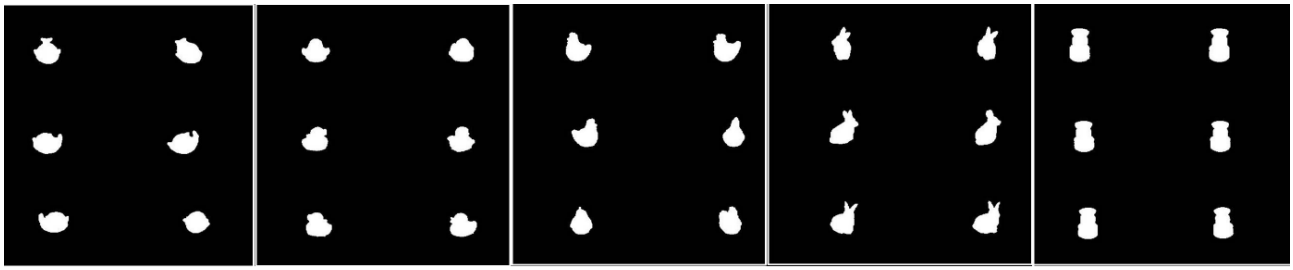

Figure 6. Sample silhouettes of the objects with unknown shapes: orange fish, duck, rooster, rabbit, and a wooden pot.

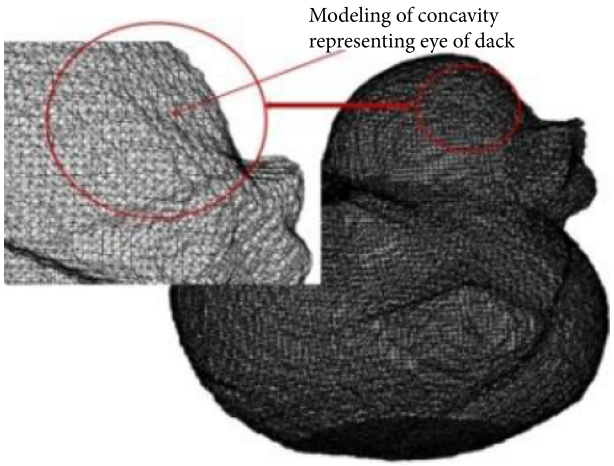

Figure 7. Surface reconstruction – surface mesh model of the 3D reconstruction of a duck.

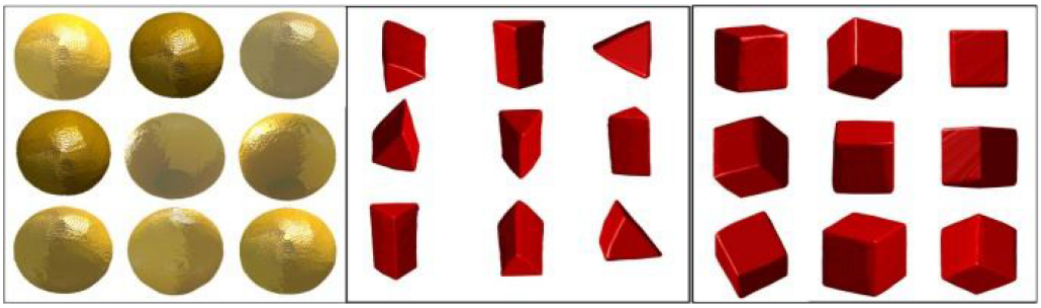

Figure 8. Sample views of the 3D reconstructed objects, sphere, prism, and cube [25], consisting of views obtained from the new viewpoints.

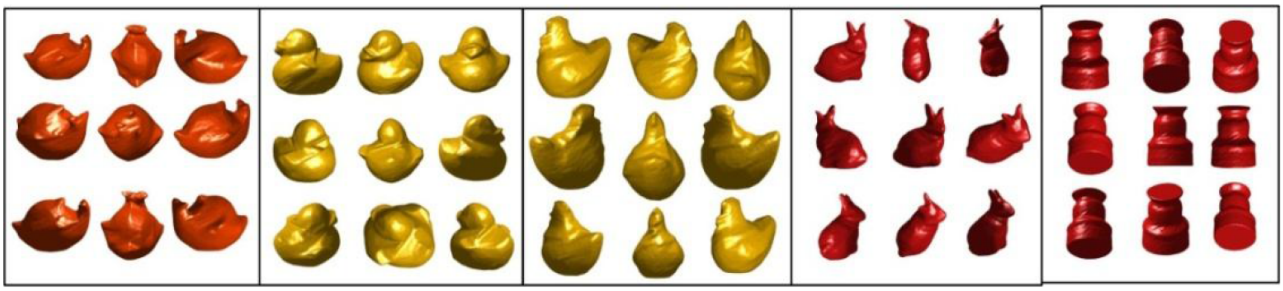

Figure 9. Sample views of the 3D reconstructed objects, orange fish, duck, rooster, rabbit, and a wooden pot, consisting of views obtained from the new viewpoints.

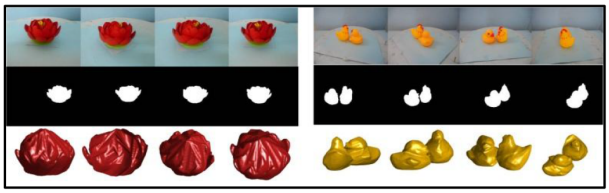

Figure 10. Left: 3D reconstruction of a fragile object – lotus, right: a scene with two objects.

Figure 11. Several views of the 3D reconstructed objects after texture mapping: orange fish, duck, rooster, rabbit, and wooden pot.

into the real-world units. It is observed that the average accuracy in the volume estimation is more than 98%. Thus, the proposed volumetric reconstruction method accurately reconstructs the objects of known as well as unknown geometrical shapes. Methods using structure from motion (SfM) face challenges while reconstructing the point cloud for the objects with fewer details. Figure 12 shows the point cloud generated for a lotus object using VisualSFM [26]. It is observed that the point cloud is sparse even after dense matching. This affects the extraction of the object surface. The dense matching approach is computationally more complex since it needs a large number of views. On the other hand, the proposed method reconstructs the object from a smaller number of views. However, the method faces challenges in reconstructing deep concavities between the petals of the lotus object.

Figure 12. Left: Point cloud obtained for the lotus object using VisualSFM [26], right: a 3D model of the lotus obtained using the proposed method.

Table 1. Comparison between the actual volume and the estimated volume of the objects with known shapes.

Object with known geometry under the 3D reconstruction	Actual volume obtained from the geometrical measurements (cm^3)	Volume measured using the water displacement method (cm^3)	Volume estimated using the proposed method (cm^3)
Sphere with diameter of 6.5 cm	143.79	144.21	144.02
Prism with height of 7.5 cm and side of triangular faces of 2.5 cm	20.29	20.34	20.88
Cube with side of 6 cm	216	216.15	216.36

Table 2. Comparison between the real-world volume and the estimated volume of the objects with unknown shapes.

Object with unknown geometry under the 3D reconstruction	Actual (real-world) volume measured using the water displacement method (cm^3)	Volume estimated using the proposed 3D reconstruction method (cm^3)
Orange fish	54.78	55.39
Duck	58.43	59.33
Rooster	47.67	48.26
Rabbit	49.65	50.16
Wooden pot	51.12	52.15

Performance of the proposed method is compared with the performance of the method using a single camera and the turntable method on the basis of computation time and reconstruction completeness [27]. The completeness percentage is the percentage of the object shape reconstructed with respect to the ground truth. Figure 13 shows the comparison between the proposed method and the single-camera method on the basis of completeness [27]. It is observed that the use of homography estimates along with the object silhouettes for computation of the voxel occupancy in a multiple-camera environment improves the accuracy of the proposed method. Table 3 presents a similar comparison on the basis of the computation time. It is observed that the computation time in the proposed multiple-camera method is longer than in the single-camera method. The computation time is affected mainly by the time required for obtaining the final voxel volume. Table 4 shows the comparison of the proposed method with the existing multiple-view reconstruction methods on the basis of the reconstruction completeness and accuracy. The number in the third column describes the accuracy of reconstruction. The number #. ## indicates that more than 90% of the surface points of the reconstructed object are within #.## (in mm) distance from the corresponding points in the ground truth. Results show that the reconstruction quality of the proposed method is on par with the methods mentioned in Table 4.

The performance of the method is further analyzed by a varying number of views and voxel resolutions. The effect of the increase in number of views on the reconstruction completeness and computation time is shown in Figure 14. It is observed that the increase in the number of views increases the quality of the reconstruction at the cost of the increase in computation time. As the quality of reconstruction is very poor for a number of views less than 9, the number of views is increased from 9 to 36 in three steps. Figure 15 shows the effect of the increase in the number of voxels on the reconstruction completeness and computation time. The number of voxels is increased in powers of 10 from 10^2 to 10^6. The logarithm of the number of voxels and corresponding

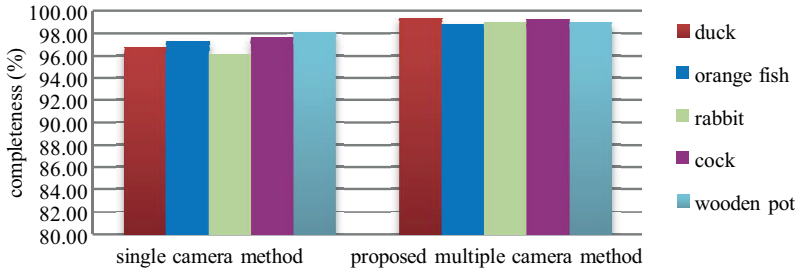

Figure 13. Comparison between the performance of the proposed method and the single-camera method [27] on the basis of completeness.

Table 3. Comparison between the proposed method and the single-camera method [27] on the basis of the computation time.

Object	Computation time (s)	
	Single camera method	Proposed method
Orange fish	70.19	117.31
Duck	61.96	99.67
Rooster	63.18	105.30
Rabbit	105.33	149.57
Wooden pot	81.41	137.58

Table 4. Comparison between the performance of the proposed method and other multiple-view methods evaluated on the basis of the reconstruction completeness.

Reconstruction method	Completeness (%)	Accuracy (in mm)
Vogiatzis et al. [28]	96.20	0.49
Kolmogorov and Zabih [29]	99.50	2.85
Hernandez and Schmitt [30]	99.50	0.45
Goesele et al. [31]	86.20	0.46
Furukawa and Ponce [32]	99.20	0.37
Vu et al. [33]	99.70	0.53
Proposed approach	99.33	0.43

Note: % completeness is the percentage of the object shape reconstructed with respect to the ground truth. The number #.## in the third column indicates that more than 90% of the surface points of the reconstructed object are within #.## (in mm) distance from the corresponding points in the ground truth.

values of completeness and computation time are plotted in respective graphs. The quality of the reconstruction is very poor for numbers of voxels less than 10,000. A more accurate reconstruction can be achieved with higher voxel resolution. However, it will increase the computation time in manifold.

5. Conclusion

The paper presents a simple and accurate multiple-view volumetric 3D reconstruction method using the integrated approach based on homography estimation and voxel mapping. The proposed 3D reconstruction

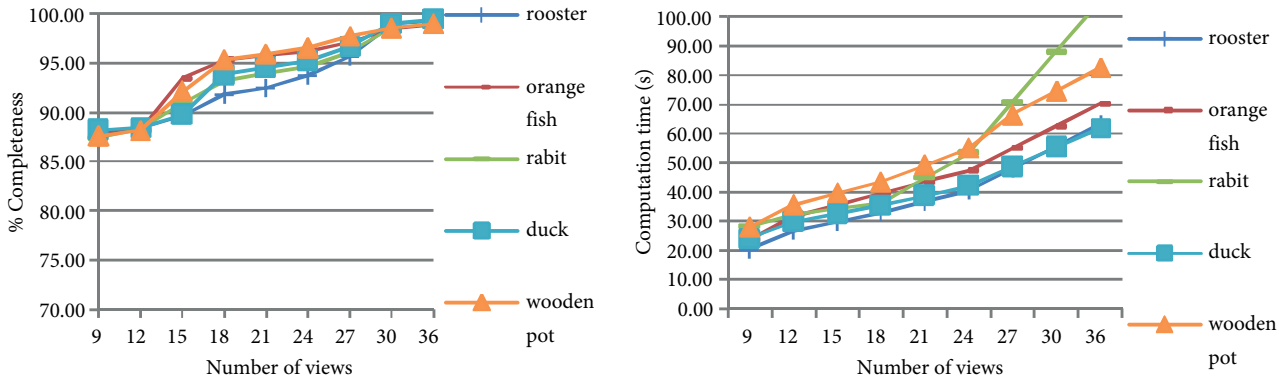

Figure 14. Left: Effect of the increase in the number of views on the reconstruction completeness. Right: Effect of the increase in the number of views on the computation time.

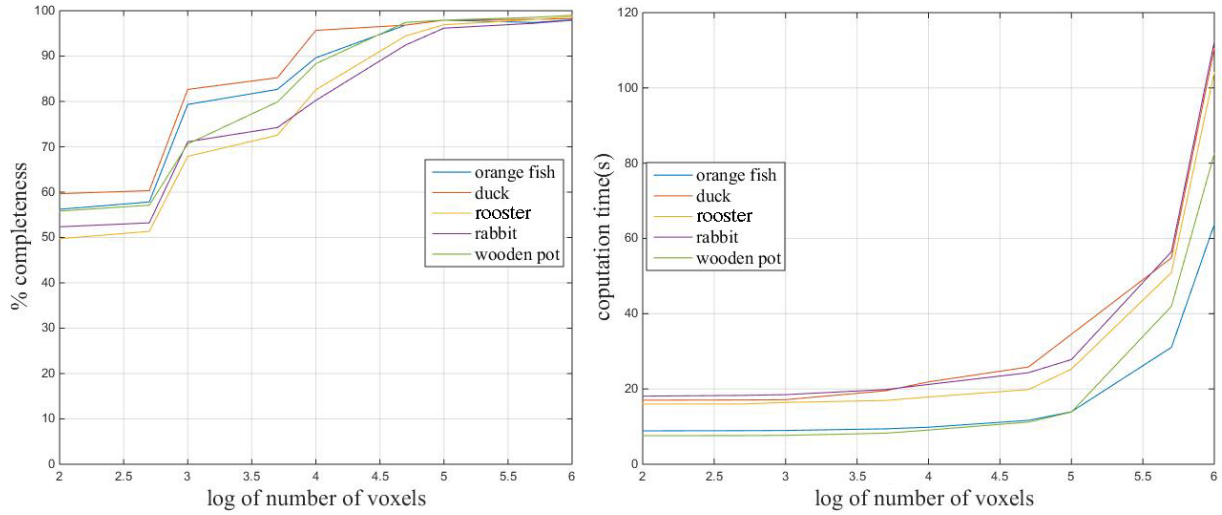

Figure 15. Left: Effect of the increase in the number of voxels on the reconstruction completeness. Right: Effect of the increase in the number of voxels on the computation time.

method combines homography estimation and the voxel mapping approach for improving the accuracy of the 3D reconstruction. The details of the reconstruction pipeline are described step by step with the results. The experimental results show that the method reconstructs objects of known and unknown shape, fragile objects, and complex scenes with multiple objects efficiently. The results prove that the reconstruction quality is good. The use of homography along with voxel mapping in a multiple-camera environment brings out more details of the object for improving the quality of reconstruction. The results show that the proposed method offers better quality of reconstruction with multiple cameras and a turntable. Cameras placed at different elevations help in bringing out more details of the object. The performance of the proposed method is on par with the existing multiple-view 3D reconstruction methods. The proposed method has better accuracy in volume estimation. Hence, the method can be effectively used in industries for volumetric measurements and analysis. The performance of the method in modeling deep concavities can be further improved.

Acknowledgment

The authors wish to thank the All India Council for Technical Education (AICTE), New Delhi, for funding the proposed work (Sanction No. 8023/RID/RPS-160 (PVT)/2011-12).

References

[1] Gallo A, Muzzupappa M, Bruno F. 3D reconstruction of small sized objects from a sequence of multi-focused images. J Cult Herit 2014; 2: 173-182.

[2] Seo Y, Choi S, Kim H, Hong K. Where are the ball and players? Soccer game analysis with color-based tracking and image mosaic. In: Image Analysis and Processing Conference; 17 September 1997; Berlin, Germany. pp. 196-203.

[3] Iwase S, Saito H. Tracking soccer players based on homography among multiple views. In: Visual Communications and Image Processing; June 2003; Lugano, Switzerland. pp. 283-292.

[4] Labayen M, Olaizola IG, Aginako N, Florez J. Accurate ball trajectory tracking and 3D visualization for computer-assisted sports broadcast. Multimed Tools Appl 2014; 73: 1819-1842.

[5] Chang J, Lee K, Lee S. Multiview normal field integration using level set methods. In: IEEE 2007 Computer Vision and Pattern Recognition Conference; 18–23 June 2007; Minneapolis, MN, USA. New York, NY, USA: IEEE. pp. 1-8.

[6] Seitz S, Curless B, Diebel J, Scharstein D, Szeliski R. A comparison and evaluation of multi-view stereo reconstruction algorithms. In: IEEE 2006 Computer Vision and Pattern Recognition; 17–22 June 2006. New York, NY, USA: IEEE. pp. 519-528.

[7] Fuhrmann S, Langguth F, Moehrle N, Waechter M, Goesele M. MVE – An image-based reconstruction environment. Comput Graph 2015; 53: 44-53.

[8] Jang I, Cho J, Lee K. 3D human modeling from a single depth image dealing with self-occlusion. Multimed Tools and Appl 2012; 58: 267-288.

[9] Shen S. Accurate multiple view 3D reconstruction using patch-based stereo for large-scale scenes. IEEE T Image Process 2013; 22: 1901-1914.

[10] Bottino A, Laurentini A. Introducing a new problem: shape-from-silhouette when the relative positions of the viewpoints is unknown. IEEE T Pattern Anal 2003; 25: 1484-1492.

[11] Srinivasan P, Liang P, Hackwood S. Computational geometric methods in volumetric intersection for 3D reconstruction. Pattern Recogn 1990; 23: 843-857.

[12] Frank S, Sturm J, Cremers D. Volumetric 3D mapping in real-time on a CPU. In: IEEE Robotics and Automation Conference; 31 May–7 June 2014; Hong Kong. New York, NY, USA: IEEE. pp. 2021-2028.

[13] Tabb A. Shape from silhouette probability maps: reconstruction of thin objects in presence of silhouette extraction and calibration error. In: IEEE 2013 Computer Vision and Pattern Recognition; 8–10 June 2013; Boston, MA, USA. New York, NY, USA: IEEE. pp. 161-168.

[14] Chen J, Bautembach D, Izadi S. Scalable real-time volumetric surface reconstruction. ACM T Graphic 2013; 32: 113: 1-16.

[15] Wu C, Liu Y, Dai Q, Wilburn B. Fusing multiview and photometric stereo for 3D reconstruction under uncalibrated illumination. IEEE T Vis Comput Gr 2011; 17: 1082-1095.

[16] Wu Z, Song S, Khosla A, Yu F, Zhang L, Tang X, Xiao J. 3D shapenets: A deep representation for volumetric shapes. In: IEEE 2015 Conference on Computer Vision and Pattern Recognition; 7–12 June 2015; Boston, MA, USA. New York, NY, USA: IEEE. pp. 1912-1920.

[17] Yan X, Yang J, Yumer E, Guo Y, Lee H. Perspective transformer nets: Learning single-view 3d object reconstruction without 3d supervision. In: Advances in Neural Information Processing Systems; 5–10 December 2016; Spain. pp. 1696-1704.

[18] Hane C, Zach C, Cohen A, Angst R, Pollefeys M. Joint 3D scene reconstruction and class segmentation. In: IEEE 2013 Computer Vision and Pattern Recognition; 25–27 June 2013; Portland, OR, USA. New York, NY, USA: IEEE. pp. 97-104.

[19] Zhao C, Xiao W. A parallelizing and improving on voxel colouring technology for 3D reconstruction based on images. In: Proceedings of the IEEE International Conference on Mechatronics and Automation; 29 July–1 August 2005; Canada. New York, NY, USA: IEEE. pp. 2185-2189.

[20] Feldmann T, Brand K, Worner A. Enhancing voxel carving by capture volume calculations. In: Proceedings of the IEEE International Conference on Image Processing; 26–29 September 2010; Hong Kong. New York, NY, USA: IEEE. pp. 4053-4056.

[21] Tsai R. A versatile camera calibration technique for high accuracy 3D machine vision metrology using off-the-shelf TV cameras and lens. IEEE J Robot Autom 1981; 24: 381-395.

[22] Bouguet JY. Camera Calibration Toolbox for MATLAB. Natick, MA, USA: MathWorks, 2006.

[23] Zhang Z. A flexible new technique for camera calibration. IEEE T Pattern Anal 2000; 22: 1330-1334.

[24] Turner E, Zakhor A. Watertight planar surface meshing of indoor point-clouds with voxel carving. In: IEEE 3DTV Conference; 7–8 October 2013; United Kingdom. New York, NY, USA: IEEE. pp. 41-48.

[25] Jadhav T, Singh K, Abhyankar A. A review and comparison of multi-view 3D reconstruction methods. J Eng Res-Kuwait 2017; 5: 50-72.

[26] Wu C. Towards linear-time incremental structure from motion. In: IEEE 2013 3D Vison Conference; 29 June–1 July 2013; Seattle, WA, USA. New York, NY, USA: IEEE. pp. 127-134.

[27] Fremont V, Chellali R. Turntable-based 3D object reconstruction. In: IEEE 2004 Cybernetics and Intelligent Systems Conference; 1–3 December 2004; Singapore. New York, NY, USA: IEEE. pp. 1277-1282.

[28] Vogiatzis G, Torr P, Cipolla R. Multi-view stereo via volumetric graph-cuts. In: IEEE 2005 Computer Vision and Pattern Recognition; 20–26 June 2005; San Diego, CA, USA. New York, NY, USA: IEEE. pp. 391-398.

[29] Kolmogorov V, Zabih R. Multi-camera scene reconstruction via graph cuts. In: European Conference on Computer Vision; 28–31 May 2002; Copenhagen, Denmark. pp. 82-96.

[30] Hernandez C, Schmitt F. Silhouette and stereo fusion for 3D object modeling. Comput Vis Image Und 2004; 96: 367-392.

[31] Goesele M, Curless B, Seitz SM. Multi-view stereo revisited. In: IEEE Computer Vision and Pattern Recognition; 17–22 June 2006. New York, NY, USA: IEEE. pp. 2402-2409.

[32] Furukawa Y, Ponce J. Accurate, dense, and robust multiview stereopsis. IEEE T Pattern Anal 2010; 32: 1362-1376.

[33] Vu HH, Labatut P, Pons JP, Keriven R. High accuracy and visibility-consistent dense multiview stereo. IEEE T Pattern Anal 2012; 34: 889-901.