3D Surface Reconstruction from 2D Multi-view Images using Voxel Mapping

[1]Tushar Jadhav, [2]Kulbir Singh, [3]Aditya Abhyankar

[1] Research scholar, [2] Professor, [3]Dean
[1] Department of Electronics & Telecommunication,Thapar University, Patiala, India
[2]Electronics & Communication Engineering Department, Thapar University, Patiala, India
[3] Department of Technology, Savitribai Phule Pune University, Pune, India
[1]tushar.jadhav@viit.ac.in, [2]ksingh@thapar.edu,[3]aditya.abhyankar@unipune.in

Abstract- **Today, there are various applications such as virtual reality, gaming, surveillance and biometrics, where 3D surface estimation and reconstruction becomes vital. Accurate 3D reconstruction of the object having complex shape is still a challenging problem in computer vision. This work is aimed towards building the computational system, which uses a probabilistic approach for computation of voxel occupancy and reconstruction of 3D shape or surface from given set of multi-view images captured from arbitrarily-distributed cameras. This is a two-step problem consisting of geometric volume intersection and refinement of this volume with photo-consistency criteria. Both the steps are implemented in order to apply the algorithm on five sets of multi-view images of the objects; Beethoven, bird, bunny, head and pig. The results obtained show the adequacy of the approach for the problem of 3D surface estimation and reconstruction. The proposed algorithm is also tested for time complexity and completeness of reconstruction for different number of cameras.**

I. INTRODUCTION

3D surface reconstruction from multi-view 2D images is an important problem in the field of computer vision. There are several attempts made by the researchers to reconstruct the 3D shape of an arbitrary and unknown shaped object from number of images captured from arbitrarily-distributed known viewpoints [1-3]. The 3D information of visible point in the object can be determined using triangulation as shown in figure1.

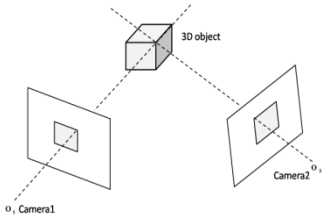

Fig 1: 3D depth from 2 views forming triangulation

Complete shape of the object can be determined using more number of cameras covering almost all the view angles such that each point from surface of the object is visible from at least N viewpoints out of all T view images, where $N \leq T$. Before reconstructing the 3D shape, it is important to have object segmented from the background in all the view images. However, segmenting the object from 2D image is an independent area in computer vision and it is out of scope of the work presented in this paper. Thus, it is reasonable to assume that the segmentation of the object from all the view images is done and segmented images are available in the form of binary silhouette images. 3D reconstruction of the object can be obtained by using back projection of silhouette objects from multi-view images into 3D space, composed of voxels. The geometric intersection of these back projection rays from the silhouette binary images forms the 3D object as shown in figure 2a and 2b.

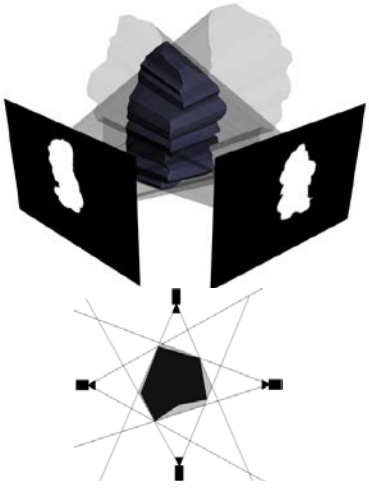

Fig 2: a) Silhouette back projection intersection forming 3D space of the object b) ray-lines from each camera forming an object's 3D space [1]

Since, 3D surface reconstruction based on geometric intersection employs only silhouette images, which are binary, and not the texture and colour information from multi-view images, voxels occupied by the reconstruction may not satisfy the photo consistency requirements. This leads to poor reconstruction of the object. Thus, the problem of 3D shape reconstruction is a two step problem consisting of computation of voxel occupancy for the scene where object is present and application of voxel mapping for removal of the voxels which are not photo-consistent.

Rest of the paper is organized as follows. Section II focuses on the related work done by other researchers in the field of 3D reconstruction. The methods for 3D reconstruction can be classified on the basis of approach used for the reconstruction as volumetric reconstruction methods and image based reconstruction methods [3]. The approach proposed in the paper is based on volumetric reconstruction and it is presented in section III. Section IV presents the results of the experimentation using the proposed approach. Section V presents the discussion and conclusion followed by the reference section.

II. Related Work

Early work in the field of 3D reconstruction based on volumetric modelling used primarily the geometric knowledge. Assuming the given 3D scene volume in voxels, the occupancy of each voxel is to be calculated. Later, to remove the spurious voxels, threshold value is applied to the occupancy of each voxel to determine whether it is an empty or occupied (filled) voxel. The earliest work in this area based on volumetric modelling using geometric information to compute a visual hull is presented by Martin and Agrawal [4]. The similar work is later extended by modelling the object using contours on successive parallel planes in [5]. Later, Laurentini [6] addressed the problem of computing parts of a non-convex object which are important for silhouette based image understanding. Another important work presented in [7] uses an energy function whose global minimum corresponds to the accurate 3D surface obtained by graph cuts in order to minimise the error caused by silhouette intersection method.

Despite being simple, volume intersection methods can be supplemented by checking the consistency of pixel appearance across the multi-view images. The use of intensity and colour information in silhouette based 3D reconstruction methods is called as ‘Shape from photo-consistency’. The photo-consistency constraint of 3D point defined as its corresponding projected points in multi-view images, where it is visible, has same intensity and the colour. The earlier work [8] dealing in photo-consistency introduces the concept of ‘true multi-image’ and the method based on it, is called as ‘Space-Sweeping Algorithm’. The approach described in this paper back-project the features obtained from each multi-view image and then takes a decision about 3D locations. The areas of the 3D space (scene) where several image features viewing rays intersect each other are likely to be the 3D locations of the observed scene features. The improved version of the ‘space-sweeping algorithm’ which accounts for occlusions is presented in [9] and it is called as ‘voxel colouring algorithm’. The algorithm is based on ordinal visibility constraint that observes the colour invariance in the corresponding pixels from all multi-view images. Another approach called, ‘space carving algorithm’ [10] uses three types of photo-consistencies to determine the correspondence pixels across all images for reconstruction of 3D shape.

The method proposed by Cheung et al. [11] is based on multiple cameras and uses background subtraction for separating the object from the background. However, this method faces segmentation related challenges. Many reconstruction methods employ photo consistency measure for checking consistency of voxels. The robustness of such measures governs the quality of 3D reconstruction. If the photo consistency is made invariant to the illumination used during acquisition of images, the reconstruction quality can be improved [12]

Reconstruction of the object surface is important step in 3D reconstruction. Many methods use signed distance fields for reconstruction of polygon meshes. Use of unsigned distance fields eliminates the need of information about the orientation of local surface. The method efficiently reconstructs the mesh for noisy data consisting holes too [13]. Similar attempt is made by Qianqian *et al*. [14] for developing a generalized mesh generation tool. The tool provides many facilities needed for generation of a mesh from multiple images.

The approach presented in [15] use multi-resolution mapping for volumetric reconstruction. The algorithm is implemented on standard CPU for real time applications. The proposed algorithm uses two step approach based on voxel mapping to reconstruct the 3D object from number of views and the silhouettes of the object.

III. The Proposed Approach

This section presents the two step algorithm for the proposed approach. The specifications of the proposed problem of 3D surface reconstruction are given as follows

- Given set of Inputs –
  - Colour/gray scale images captured from N cameras i.e. N multi-view images, I_C (m, n).
  - Silhouette binary images, each representing the object of interest from the scene and obtained from the multi-view images, B_C (m, n),
  - Camera perspective projection matrix, H_C of dimension 3 x 4 for all the cameras used to obtain multiple images of the object.
- Estimated Output-
  - Set of 3D points, V(x,y,z), which is made up of voxels of the scene having either value 0 (empty) or 1(filled/occupied) to represent 3D shape of the object.

The simple scene with a rectangular object is shown in figure 3, where the object is represented by regular $2 \times 3 \times 2$ voxels. However, in reality the object can be of any number of adjoint voxels within the scene and can have irregular shape too.

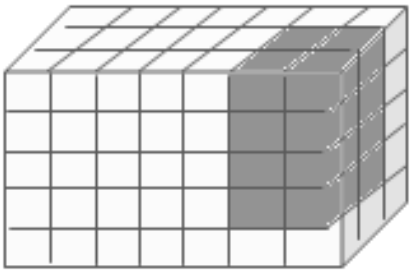

Fig 3: The scene (all voxels) and the object (shaded voxels)

The proposed approach uses geometric volume intersection method to obtain visual hull of the object. The visual hull is obtained by using voxel-occupancy assignment to each voxel from the scene. The method uses a probabilistic approach for calculation of voxel occupancies at each grid point. The method computes joint probability for each voxel in terms of silhouette probabilities, number of views and number of cameras. This voxel-occupancy metric is passed through the threshold value to identify the occupied (filled) voxels and discard the empty ones. The volume thus obtained is refined further by using photo-consistency criterion. The proposed approach employs following criterions for checking voxel consistency.

1) Occupied voxel should be present in at least M view images, which are obtained from silhouette images (weak photo-consistency criteria)

2) Occupied voxel should have minimum variance across the intensities of correspondence pixels in multi-view images.

The threshold values for maximum variance (standard deviation) and number of cameras for each dataset are given in the table 1. The following subsections discuss both the steps: geometric volume intersection and refinement of shape in detail. The flowcharts for both the steps in the algorithm are presented in figure 4 and figure 5 respectively.

TABLE 1: THRESHOLD VALUES FOR PHOTO-CONSISTENCY CRITERIONS

Sr. no	Dataset	Number of Cameras (or views) (N)	Maximum Deviation	Minimum number of Cameras
1	Beethoven	32	50	N/2
2	Bird	21	70	N/2
3	Bunny	36	80	N/2
4	Head	33	90	N/2
5	Pig	27	125	10

A. Geometric Volume Intersection

In the first step of 3D shape reconstruction, only silhouette images and corresponding perspective transformation matrices are used. After dividing the entire 3D space into voxel grid, a probabilistic approach is used for computation of voxel occupancies. The algorithm estimates joint probabilities for voxel occupancy at each grid point considering silhouettes from all the views/cameras and their probabilities. The voxels which are having occupancies greater than certain threshold are termed as opaque. Other inconsistent voxels are removed or made transparent. Thus only the voxels which are related to the object are retained to generate a voxel volume corresponding to the 3D object. This voxel mapping is supported with initial guess about the location of the object voxel. This is determined by calculating a scene range as shown in figure 4.

B. Photo-consistency based refinement

Further, in order to refine the volume obtained in geometric intersection step, the photo-consistency criterions are used. In first criteria, each voxel is tested for its visibility. If the voxel is visible in the object silhouettes of at least M view images, the voxel is considered as a consistent voxel and hence retained. In second criteria, the variance in the RGB values of the pixel from the object silhouette where voxel is back projected (not to background silhouette), are calculated. If the addition of three variances in RGB values is less than the defined threshold, then the voxel is considered to be a part of the object. The method retains only these voxels as object voxels to

generate the refined 3D shape. The algorithm of this second stage processing is shown in figure 5. Potentially, a reconstruction system proposed in this paper can be a part of many applications where 3D-information is necessary. Some applications are listed as below

- Volumetric analysis of an object in industrial inspection system.
- An architect takes pictures of a city block. For the planning of a new building it is important to have an accurate virtual model of the city block and building.
- A reconstruction system is useful for robots to navigate in an unknown environment and to build a map iteratively.
- Insertion of a synthetic object into an existing video sequence is an important task in movie making.
- A biometric system to identify people may use 3D profile of the face.
- In medical field, 3D information can be used to help doctors for guided surgeries.

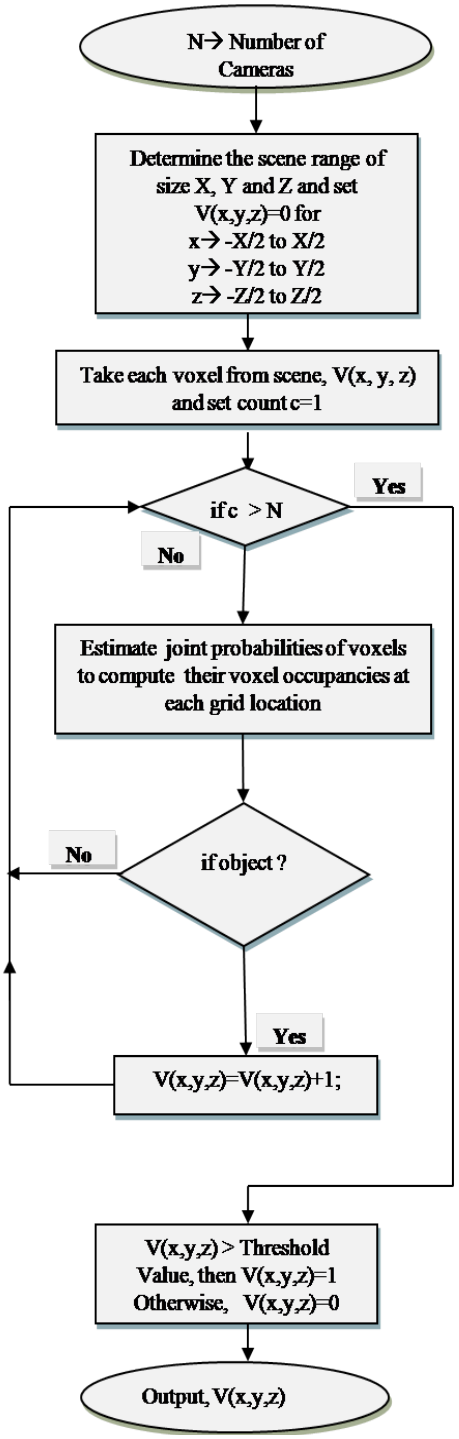

Fig 4: Algorithm for geometric volume intersection

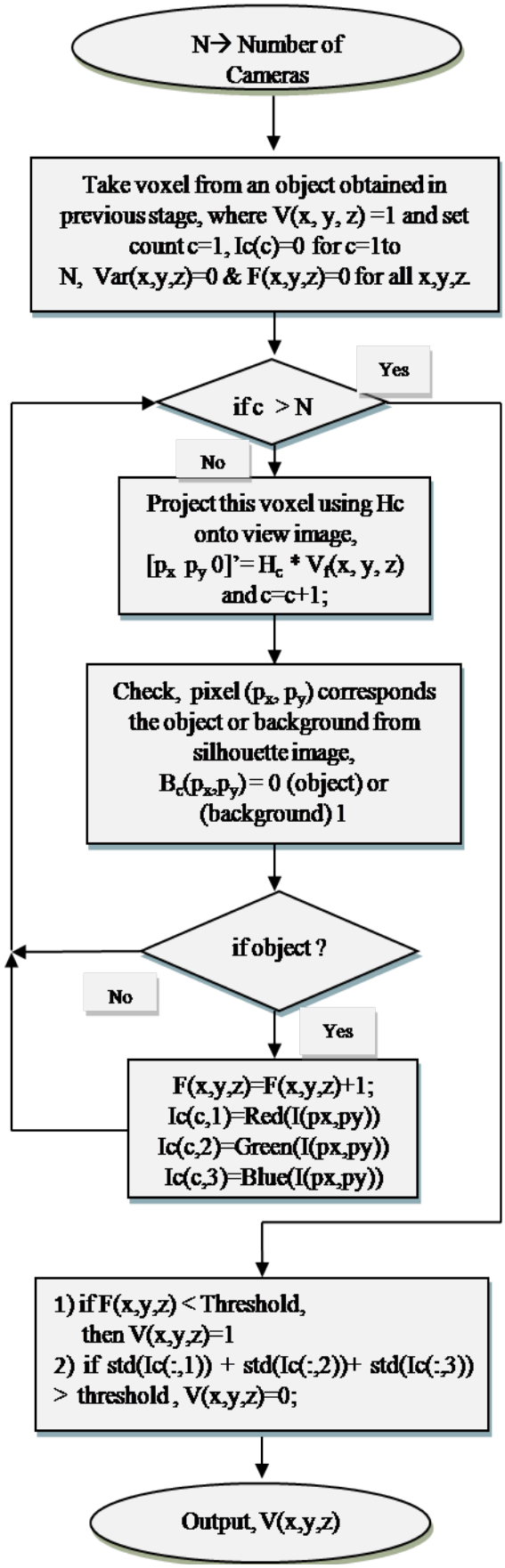

Fig 5: Algorithm for refinement with photo-consistency criterions

IV. Experimental Results

A. Dataset Description

The dataset used in this work is downloaded from the website [16]. The dataset is made publicly available. It consists of multiple images and related information for five objects - Beethoven, bird, bunny, head, and pig [17- 19]. The information about number of cameras used for imaging the object is also given in each object dataset. For each camera, following three elements are provided in the dataset:

- Colour Image, $I_c(m,n)$
- Silhouette binary image, $B_c(m,n)$
- Projection Matrix , H_c

Thus, for each dataset three folders are provided. First folder consists of colour images captured using N cameras. Second folder provides all the silhouette binary images. Last folder consists of the text files having projection matrix elements. The number of cameras used for obtaining multi-view images in each dataset is given in the table 1.

B. Experimentation

The results obtained with geometric intersection step and the refinement in the reconstruction using photo consistency criterions are presented in figure 9 to figure 18 after the reference section. The images from the multi-view image data sets for five objects- Beethoven, bird, bunny, head and pig are shown in figures 9, 11, 13, 15 and 17. The figures 10, 12, 14, 16 and 18 present the reconstruction results for corresponding objects. Each row presents different views of the reconstruction column-wise - isometric view with (30^o, 30^o) and the views from left side (90^o), front side (0^o), right side (-90^o) and back side (-180^o). First row shows the results of geometric intersection step with threshold 0.9. Second row presents results of the refinement step using the criterion defined for voxel consistency on minimum number of cameras/ silhouettes. In the last row, the reconstruction results after applying consistency criterion based on maximum variance (standard deviation) along with the criteria of presence in minimum number of cameras are shown. The second criterion is applied for further refinement of the volume and obtained views are presented in same order as mentioned above. It is evident from the results that the reconstruction obtained with refinement using both the criterions gives the 3D surface estimation very close to the actual object.

The proposed algorithm is also tested for its time complexity and completeness of reconstruction for different number of cameras. The quality of reconstruction is expressed using the parameter

'completeness'. Completeness describes the closeness of the reconstructed object with the ground truth. The silhouettes obtained from reconstructed object are compared with the ground truth and percentage error in matching is calculated. Lower the value of this error more will be the percentage of completeness. Figure 6 presents the relation between completeness of reconstruction and number of cameras. It is observed that, with increase in number of cameras the completeness of reconstruction improves to make the reconstruction closer to the actual object.

The relation between computation time required for 3D reconstruction and number of cameras is presented in figure 7. Computation time is measured as the total time required for the method to reconstruct the object by allowing only minimal services to run on the computing machine. Increase in computation time will increase the time complexity of the algorithm. The computation time increases with number of cameras. It is observed that the main contribution to the computation time is the time required for geometric intersection.

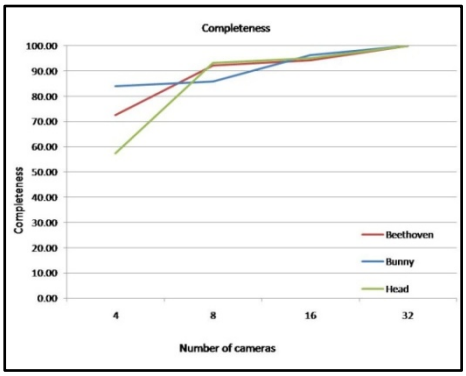

Fig 6: Completeness versus number of cameras

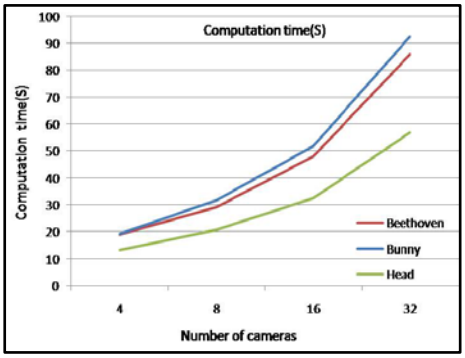

Fig 7: Computation time versus number of cameras

The reconstruction of the objects improves with photo consistency criterion. The 3D reconstruction consistent with minimum number of cameras is comparatively less close to the actual one as compared to the 3D reconstruction consistent with minimum number of cameras and pixel variance. Figure 8 presents this comparison in the form of a bar chart. Lower value of pixel variance indicates that the reconstruction is more consistent with the photo consistency criteria. Figure 19 and 20 present the 3D reconstruction of Beethoven and bunny for different number of cameras.

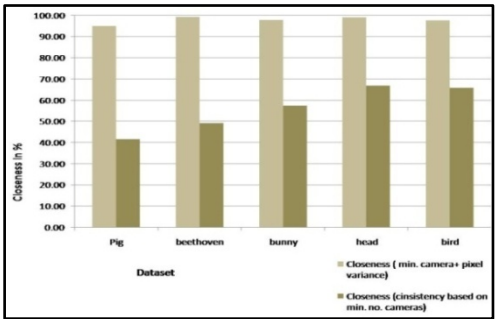

Fig 8: Closeness of reconstructed volume with expected volume for minimum number of camera and minimum number of camera + pixel variance

V. Discussion and Conclusion

This work is aimed towards building the computational system, which uses a probabilistic approach for computation of voxel occupancy and reconstruction of 3D shape or surface from given set of multi-view images captured from arbitrarily-distributed cameras. This is a two-step problem consisting of geometric volume intersection and refinement of this volume with photo-consistency criteria. The algorithm is implemented on the five sets of multi-view images for the objects; Beethoven, bird, bunny, head and pig. The result shows reconstruction with both criterions, minimum number of cameras and maximum variance, has given better approximation to the true object. The computational complexity of this algorithm is less as compared to other approaches and it is proportional to the number of views/cameras. Hence, real time implementation of the algorithm is possible. Increase in number of cameras improves the reconstruction quality that is the completeness of reconstruction whereas it also increases the computation time. The algorithm allows arbitrary placement of multiple cameras as compared to the method using turn table and single camera. Use of multiple cameras makes the algorithm suitable for 3D reconstruction of objects which cannot be rotated.

References

[1] Dyer, Charles R., "Volumetric scene reconstruction from multiple views", *Foundations of image understanding, Springer US*, pp. 469-489, 2001

[2] Seitz S., Curless B., Diebel J., Scharstein D., Szeliski R., "A comparison and evaluation of multi-view stereo reconstruction algorithms", *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 519 – 528, 2006

[3] Jadhav T., Singh K., Abhyankar A., "A review and comparison of multi-view 3D reconstruction methods", *Journal of Engineering Research (in press)*

[4] Martin W. N., Aggarwal J. K., "Volumetric descriptions of objects from multiple views," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 2, pp. 150-158, March 1983.

[5] Srinivasan P., Liang P., Hackwood S., "Computational geometric methods in volumetric intersection for 3D reconstruction", *Proceedings of IEEE International Conference on Robotics and Automation*, vol. 1, pp. 190-195, 14-19 May 1989.

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

[7] Snow D., Viola P., Zabih R., "Exact voxel occupancy with graph cuts," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 345-352, 2000.

[8] Collins R.T., "A space-sweep approach to true multi-image matching", *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '96)*, pp. 358-363, Jun 1996.

[9] Seitz S.M., Dyer C.R., "Photorealistic scene reconstruction by voxel coloring", *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1067-1073, Jun 1997.

[10] Kutulakos K.N., Seitz S.M., "A theory of shape by space carving," *Proceedings of the Seventh IEEE International Conference on* Computer Vision, vol. 1, pp. 307-314 , 1999.

[11] Cheung G.K.M., Baker S., Kanade T., "Visual hull alignment and refinement zcross time: a 3D reconstruction algorithm combining shape-from-silhouette with stereo", *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR '03)*, vol. 2, pp. 375-382, June 2003.

[12] Hornung A., Kobbelt L., "Robust and efficient photo-consistency estimation for volumetric 3d reconstruction", *Proceedings of European Conference on Computer Vision*, pp. 179-190, 2006.

[13] Hornung A., Kobbelt L., "Robust reconstruction of watertight 3 D models from non-uniformly sampled point clouds without normal information", *In Symposium on geometry processing* , pp. 41-50, June 2006.

[14] Qianqian F., Boas D.A., "Tetrahedral mesh generation from volumetric binary and grayscale images", *IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'09)*, pp. 1142-1145, 2009.

[15] Frank S., Sturm J., and Cremers D., "Volumetric 3d mapping in real-time on a CPU", *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2021-2028, 2014.

[16] http://vision.in.tum.de/data/datasets/3dreconstruction

[17] Cremers D., Kolev K., "Multi-view stereo and silhouette consistency via convex functionals over convex domains ", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, 2011.

[18] Kolev K., Klodt M., Brox T., Cremers D., "Ontinuous global optimization in multi-view 3D reconstruction" , *In International Journal of Computer Vision*, vol. 84, 2009.

[19] Kolev K., Pock T., Cremers D., " Anisotropic minimal surfaces integrating photo consistency and normal information for multi-view stereo", *Proceedings of European Conference on Computer Vision (ECCV)*, 2010

Fig 9: Multiple images of the object - Beethoven

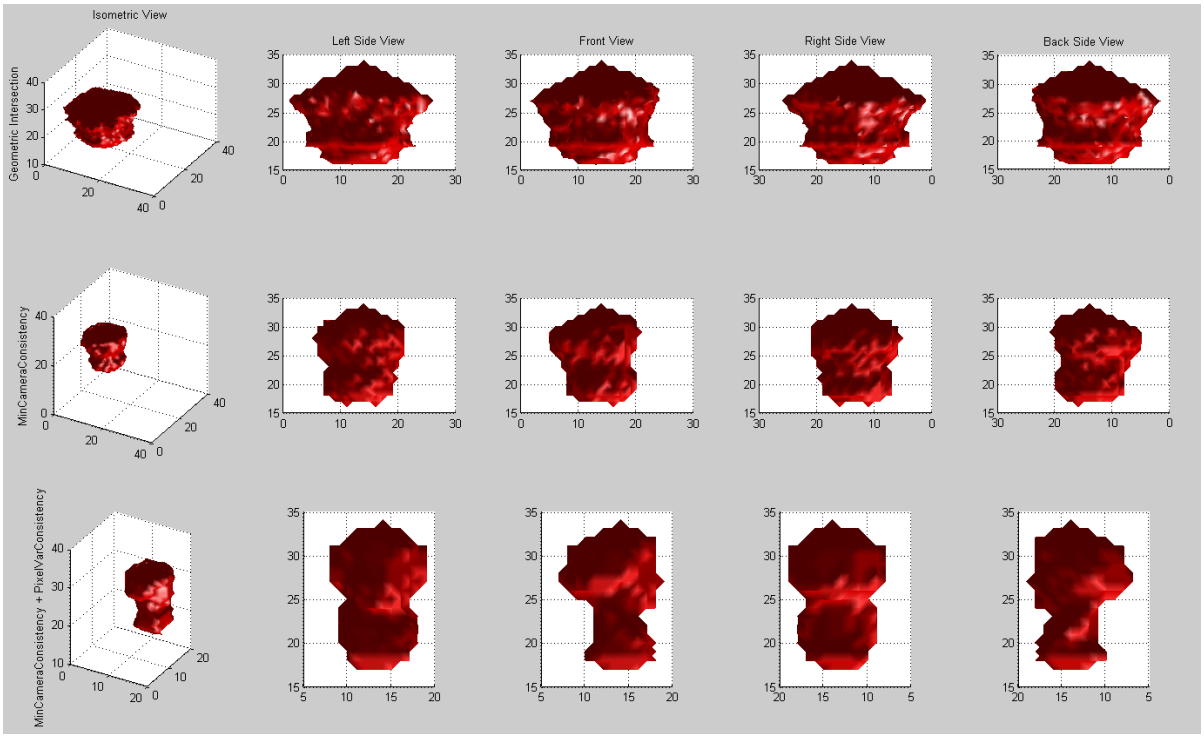

Fig 10: 3D reconstruction of the object – Beethoven: a) Row1- geometric intersection b) Row 2 and 3- refinement in reconstruction with photo consistency criterions

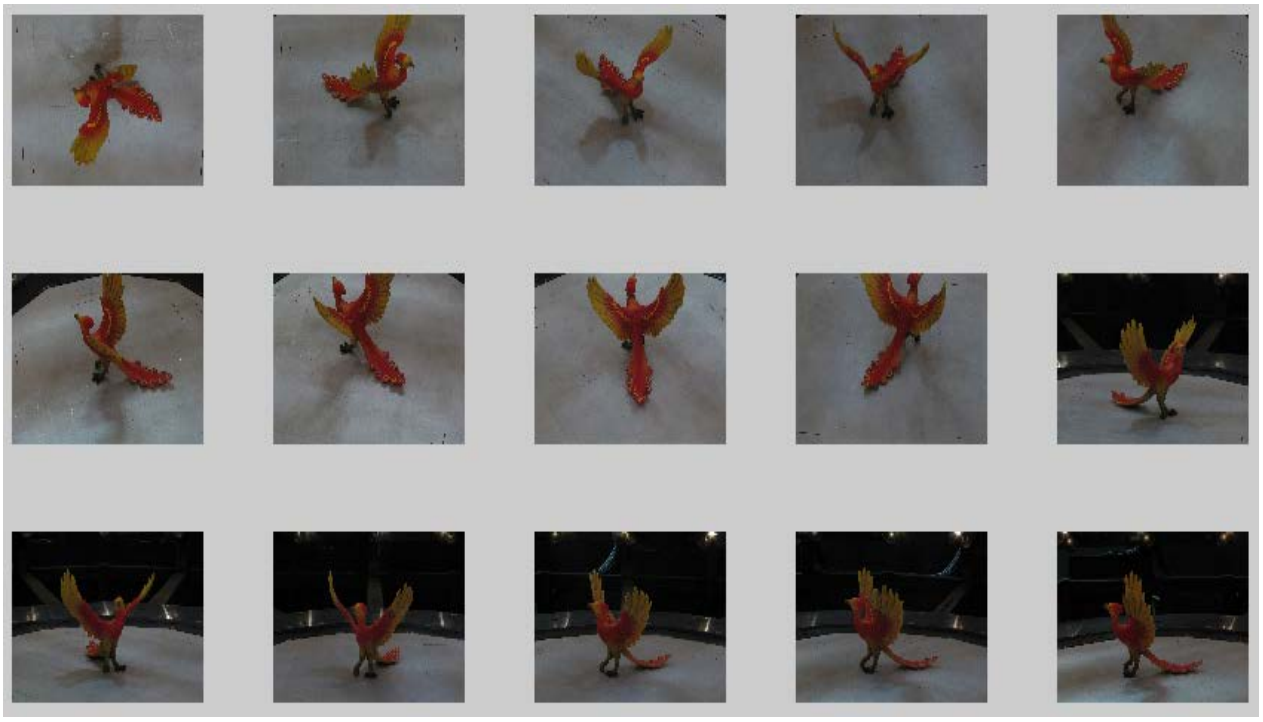

Fig 11: Multiple images of the object - bird

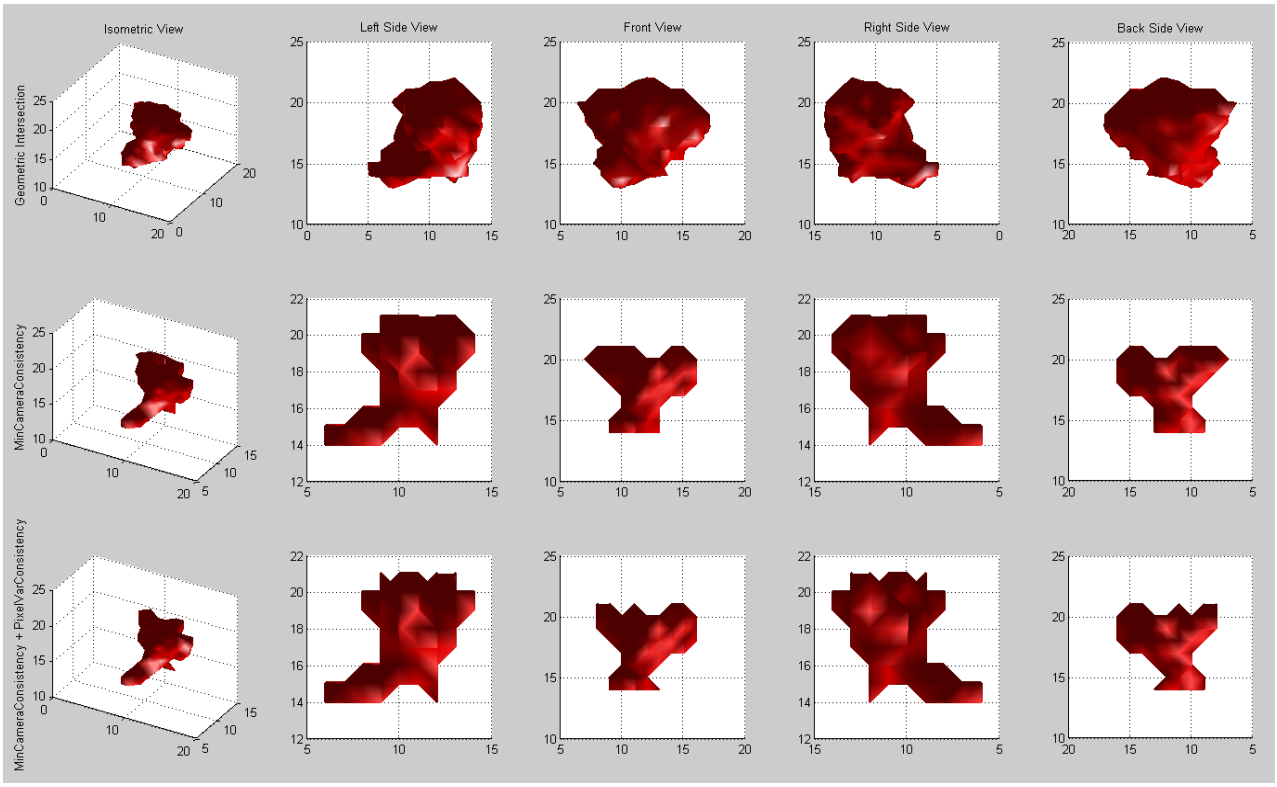

Fig 12: 3D reconstruction of the object – bird: a) Row1- geometric intersection b) Row 2 and 3- refinement in reconstruction with photo consistency criterions

Fig 13: Multiple images of the object - bunny

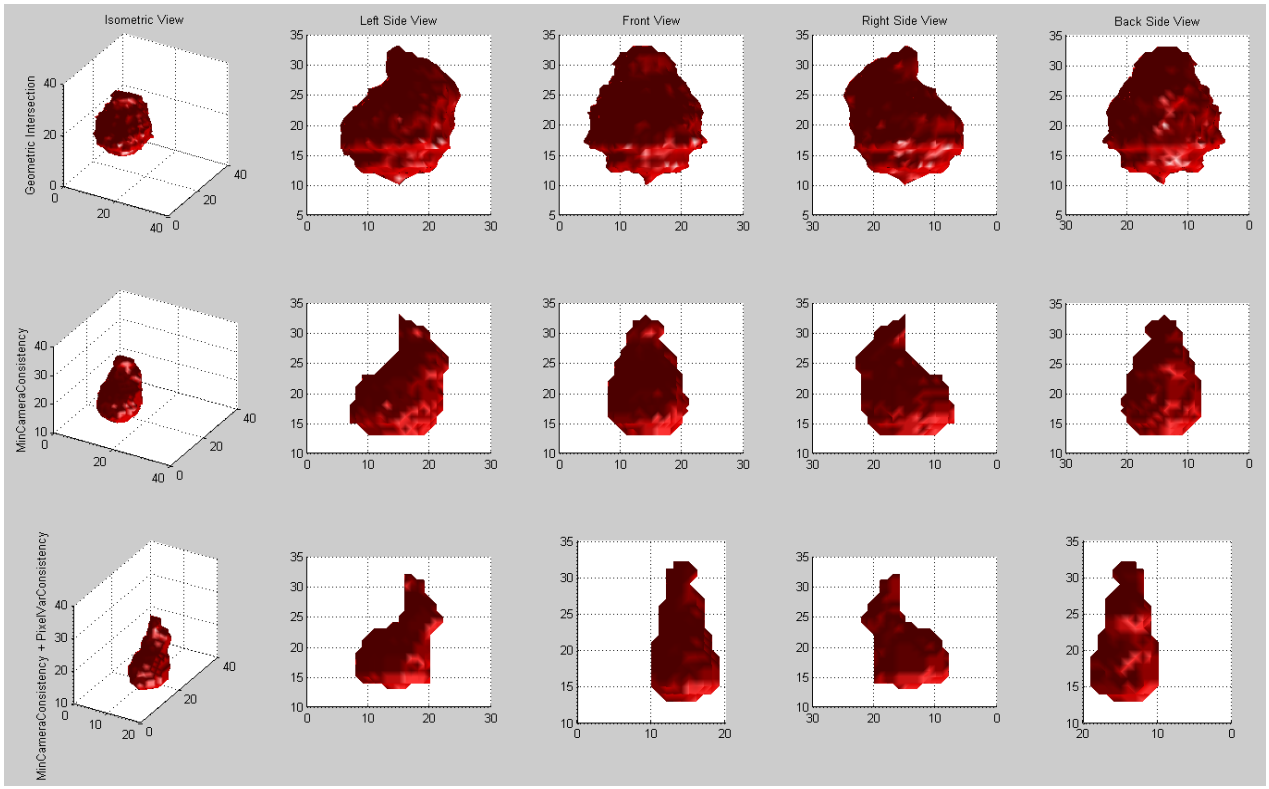

Fig 14: 3D reconstruction of the object – bunny: a) Row1- geometric intersection b) Row 2 and 3- refinement in reconstruction with photo consistency criterions

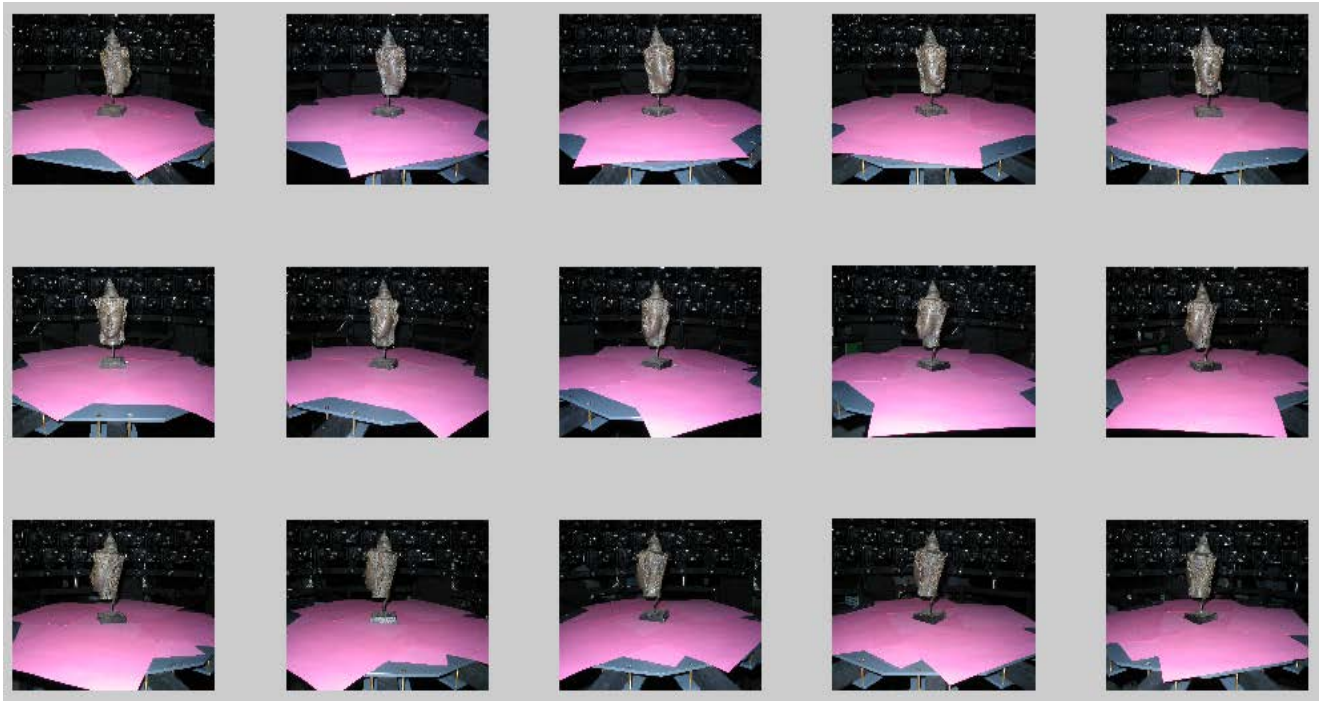

Fig 15: Multiple images of the object - head

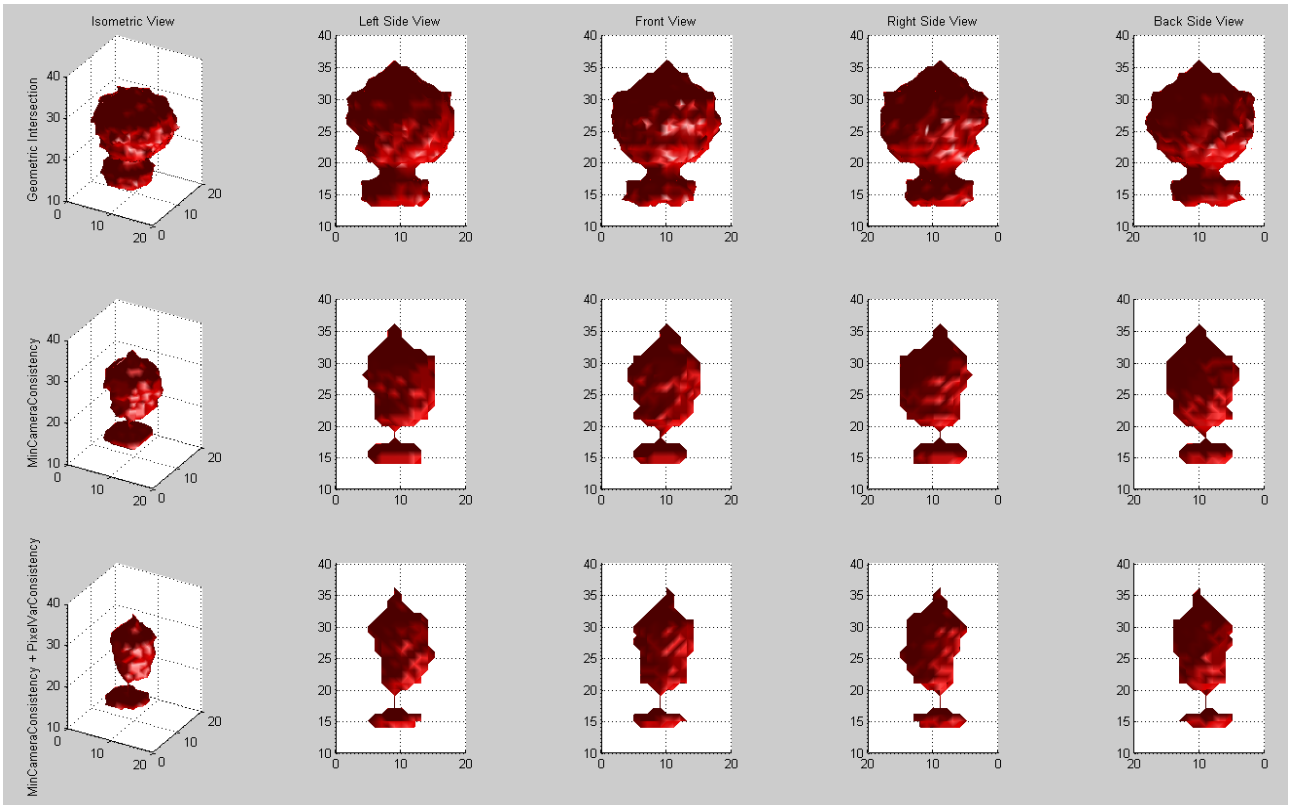

Fig16: 3D reconstruction of the object – head: a) Row1- geometric intersection b) Row 2 and 3- refinement in reconstruction with photo consistency criterions

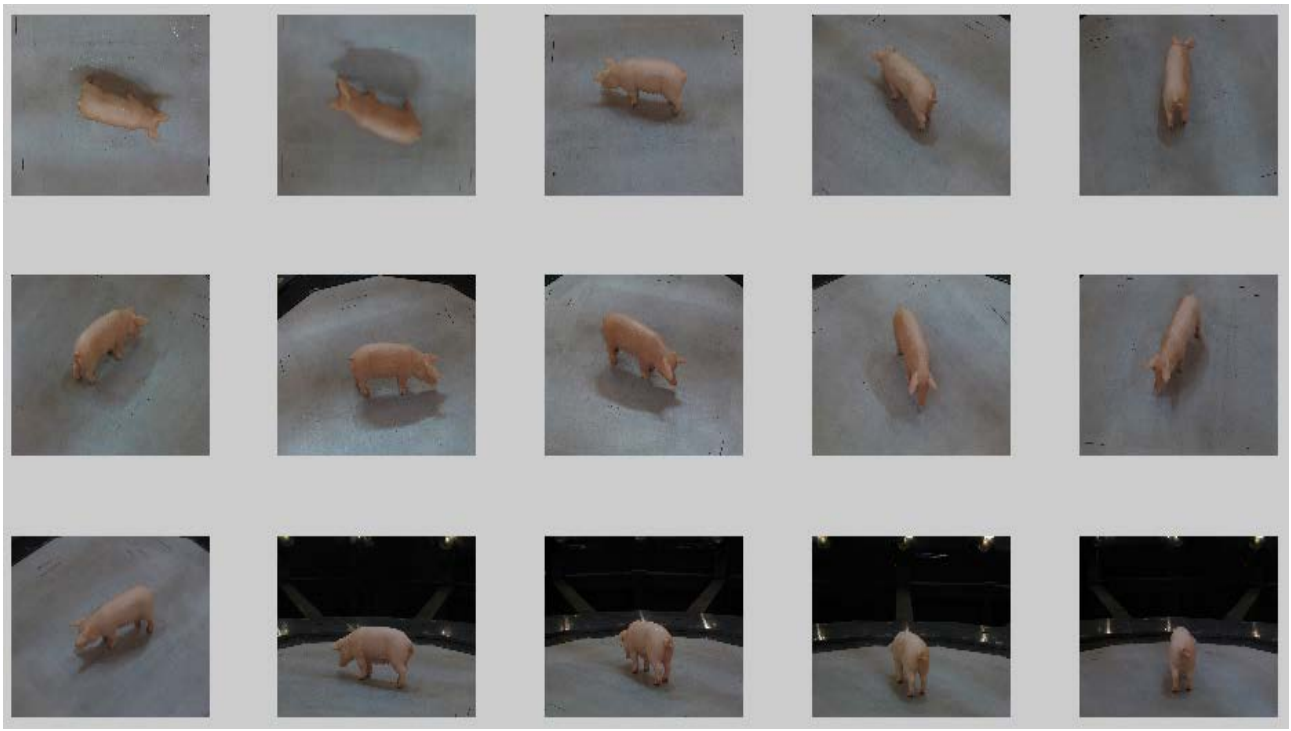

Fig 17: Multiple images of the object - pig

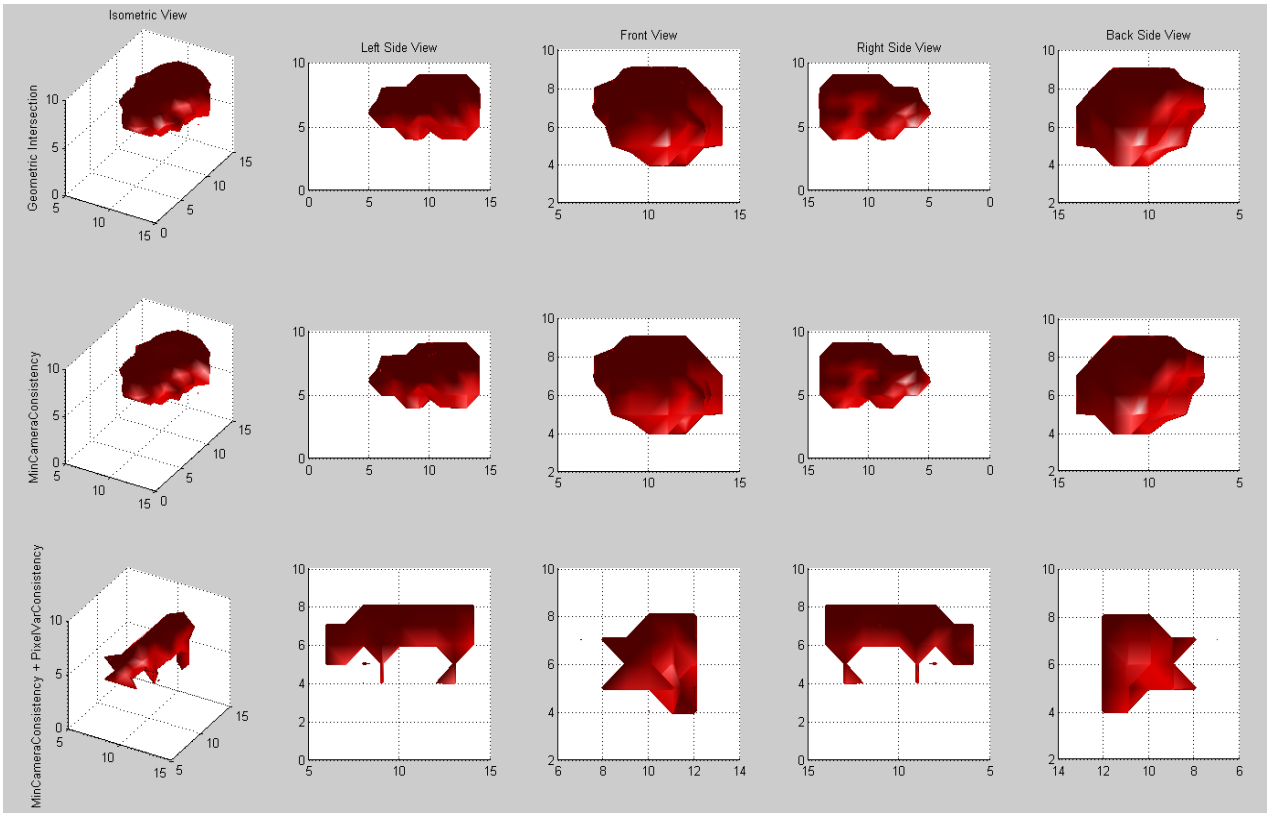

Fig 18: 3D reconstruction of the object – pig: a) Row1- geometric intersection b) Row 2 and 3- refinement in reconstruction with photo consistency criterions

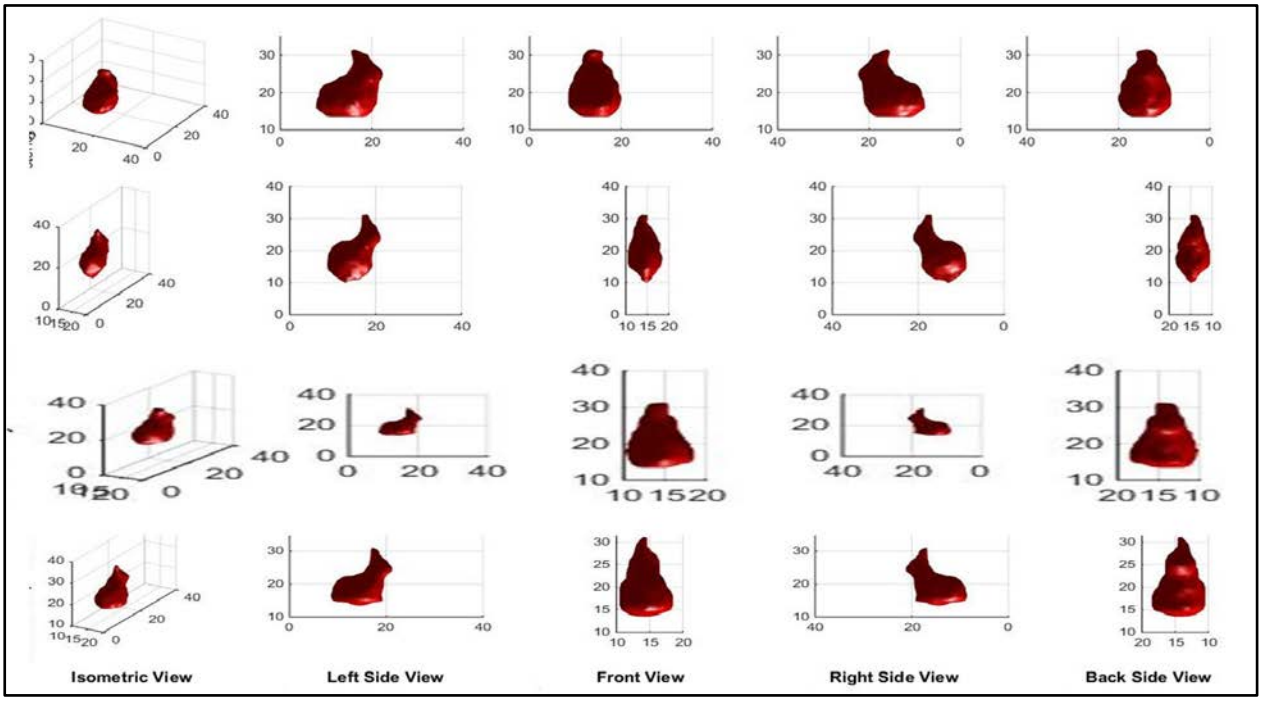

Fig 19: Effect of increase in number of cameras on quality of 3D reconstruction of the object- bunny. Reconstruction quality improves with number of cameras (first row – results with 4 cameras, second row -results with 8 cameras, third row- results with 16 cameras, last row results with 32 cameras)

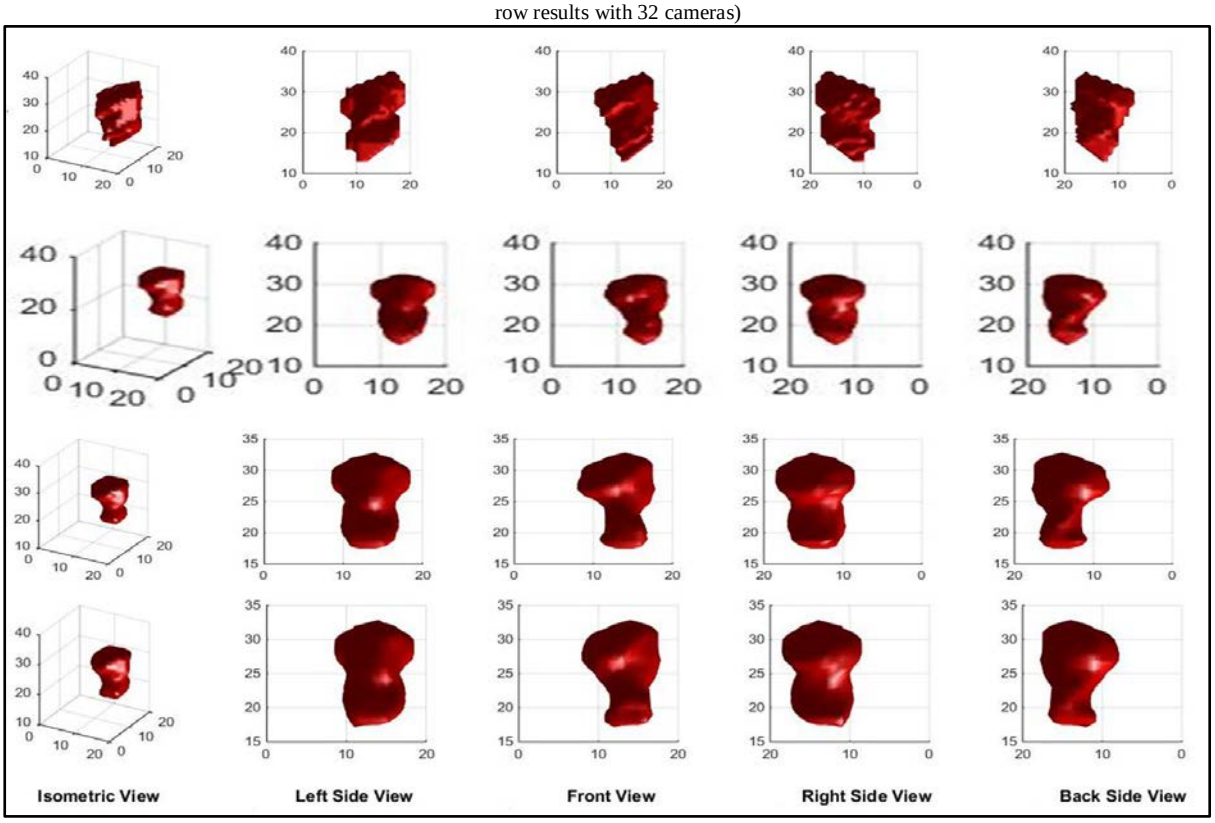

Fig 20: Effect of increase in number of cameras on quality of 3D reconstruction of object- Beethoven. Reconstruction quality improves with number of cameras (first row – results with 4 cameras, second row -results with 8 cameras, third row- results with 16 cameras, last row results with 32 cameras)