Surface Model Reconstruction of 3D Objects From Multiple Views

Vincenzo Lippiello and Fabio Ruggiero

Abstract— **A points surface reconstruction algorithm of 3D object models from multiple silhouettes is proposed in this paper. Some images of the target object are taken from a circular trajectory by a robot with a camera mounted in an eye-in-hand configuration. The silhouettes of the observed object are evaluated for each view using a blob analysis process, and from those a set of points that sample a reconstruction sphere surrounding the target object are estimated. The sphere sample points are attracted by the object center of mass using a variable step according to the distance from the silhouettes contours. For each point, the iterative process of constriction is stopped when all the back-projections of the point are within the corresponding silhouettes. Moreover, a new method based on a rough estimation of object dimension is proposed to reduce the disturbances due to projection and shadow cones. Simulations and experiments are presented to evaluate the performance of the proposed algorithm.**
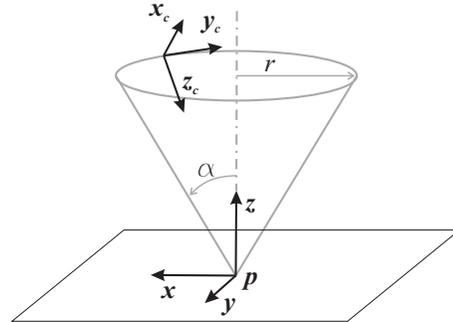
Fig. 1. Circular trajectory executed by the camera mounted on the robot hand during the image acquisition step with respect to the reference frame.

I. INTRODUCTION

For many years, reconstruction of 3D object models has been used to solve a number of applications in robotics, reverse engineering, measure and quality control, medical field.

Several methods have been proposed in the literature to achieve this goal, based on different approaches for data acquisition, such as omnidirectional cameras, single or multiple cameras mounted on robots, photos take by hands or with a tripod, etc. Some relevant differences are in how the obtained images are processed, and in particular in the algorithm used to reconstruct the object model.

One branch of algorithms is classified under the so called *volumetric scene reconstruction* approach [2]. This class of algorithms can be further divided into two main groups: the *shape from silhouettes* and the *shape from photo-consistency* approach.

The silhouette images play a crucial role in the first approach; binary images are used where the value of a pixel indicates whether or not the visual ray from the optical center intersects the object surface. The union of all visual rays defines a solid cone and the intersection of all these cones for all the images defines a volume in the scene corresponding to a first approximation of the object volume. Starting from

The authors are with PRISMA Lab, Dipartimento di Informatica e Sistemistica, Università degli Studi di Napoli Federico II, via Claudio 21, 80125, Naples, Italy `vincenzo.lippiello@unina.it`, `fabio.ruggiero@unina.it`

this one, different techniques to reconstruct the object model have been developed. One of the most used methods samples the volume in voxels [2]; in this case, the task consists of creating the voxel occupancy description corresponding to the intersection of the back-projected silhouette cones. To make the scene description more efficient, the voxels are organized in an octree representation; how to build this octree is an important research topic [4], [8], [11].

In the second approach grayscale images or color ones are used. These further information is utilized to improve the 3D model reconstruction of the target object: *voxel coloring* [10] and *space carving* [5], [14] are some of the most investigated techniques in this branch.

Another method is proposed in [13] that reconstructs the object model, starting from a surface that moves towards the object under the influence of internal forces, given by the surface itself, and external forces, given by the image data. Typically, the starting surface is a sphere. This approach may be considered as a generalization of snakes used in 2D.

A technique for computing a polyhedral representation of the *visual hull*, the set of points in the space that project inside one silhouette in every image [6], is studied in [3]. In this approach, only the silhouette contours in the images have to be visited, and the computed visual hull is quickly represented.

Furthermore, other methods rely on use of *apparent contours* [1], [9]; in these cases the reconstruction is based on the spatio-temporal analysis of deformable silhouettes.

In this paper a variant to the method proposed in [11] is presented based on the use of points instead of voxels to reduce the computational complexity. Starting from multiple 2D silhouettes of the 3D object, a sphere centered at the object center of mass is estimated, which contains the object. This sphere is first sampled using points. Each sampled point
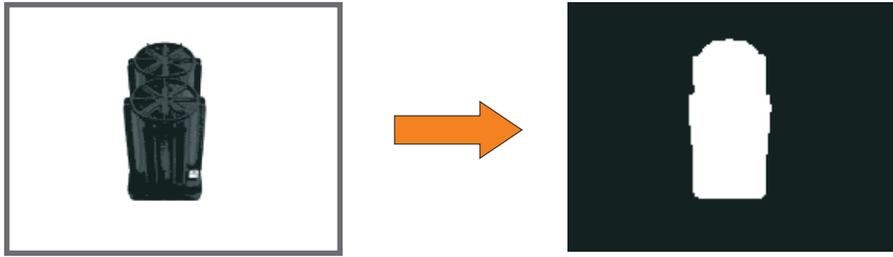
Fig. 2.   Example of the image processing employed to extract the silhouette of the observed object.

is forced to approach the estimated object center of mass until the visual hull is intersected; then the point is kept fixed, eventually after a further local refining. The process ends when all the sample points have reached the object surface; the resulting fixed points build the object reconstruction.

Moreover, a new method to reduce the negative influence of shadow and projection cone of the object model reconstruction based on an estimation of object dimensions is also proposed, based on the back-projection of the evolution axis used to generate the acquisition trajectory followed by the robot.

One of the aims of the proposed method is to achieve a sensible reduction of the computational complexity, while accepting a limited reduction of the reconstruction accuracy with respect to other approaches, to allow a real-time object model reconstruction suitable, e.g., for robotic hand grasp planning [15], [16].

The hypothesis made in this work if the use of a calibrated camera mounted on a robot in an eye-in-hand configuration, and the availability of images taken around the object from a known circular path. However, where necessary, these limitations may be overcome using the method proposed in [12]. The image elaborations are performed with classic blob analysis techniques.

The outline of the paper follows the process step indicated in [7] where a further approach to the problem, called *marching cubes*, is presented. In Section II, it is explained how the images are acquired, and the main steps for the images processing. Then in Section III the proposed algorithm is presented with the description of the steps that lead towards the points reconstruction of the object model, while in Section IV a method is described which, with a rough estimation of object dimensions, can overcome the disturbances due to projection and shadow cones. Finally, in Section V some simulations and practical experiments are presented to validate the proposed method.

II. IMAGE ACQUISITION AND PROCESSING

To acquire the images necessary to reconstruct the object model, a camera mounted on a robot in an eye-in-hand configuration is considered. The robot performs a circular trajectory surrounding the object. The use of a calibrated robotic system allows views to be taken from a circular path measuring with high precision the position of the camera
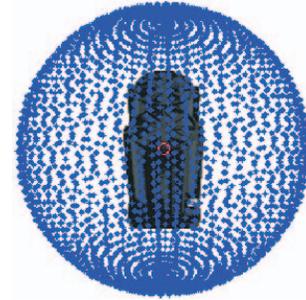

Fig. 3.   Sampled reconstruction sphere that surrounds the object.

frame of each click with respect to a common reference frame, i.e. the robot reference frame, as in Fig. 1.

The circular camera trajectory is generated automatically assigning the number of images to acquire N, a rough estimation of the object position $\mathbf{p}$, the revolution axis $\mathbf{z}$ of the observation cone, the observation angle α, and the radius r of the circular path (or equivalently the observation height), as shown in Fig. 1.

Once these N images have been acquired on this circular trajectory, the images processing step can start. First of all, the silhouettes of the images need to be determined with the use of the blob analysis. To obtain this result, the image enhancement can be considered both in spatial and in frequency domain, with the goal to reduce noise and disturbances such as the presence of shadow in the views. In this paper a simple binarization process, with a self-tuned threshold, has been employed. After this process a binary dilation and erosion may be necessary to reduce the effects of the image noise and of a imperfect illumination.

Figure 2 shows a target object before the images processing step and the resulting silhouette after the blob analysis and the image enhancement. Obviously, this process must be performed for all the acquired images.

III. MODEL ESTIMATION

For each of the N silhouettes that have been extracted, the centroid of the corresponding blob is evaluated. Using a triangulation method a rough center of mass of the object can be estimated. Then, for each image, the radius r_i of the minimum circle in which the silhouette is surely inside is evaluated. The number

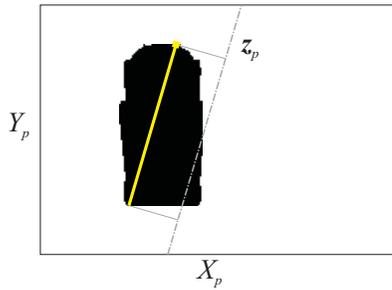$$r_s = \max \ r_i + \epsilon \quad i = 1, \ldots, N$$

Fig. 4. The dot-dashed line represents the back-projection $\mathbf{z}_p$ of the axis $\mathbf{z}$ on an image plane; the yellow line represents the maximum chord of the silhouette parallel to $\mathbf{z_p}$.

is the radius of the sphere that in 3D space is inside the real object, with ϵ an enlarging coefficient. A reconstruction sphere with radius r_s, center corresponding to the estimated object center of mass, and sampling points density d can be built as shown in Fig. 3.

Using a dynamic step depending on the distance from the object surface, the sphere radius is reduced. When a sample point of the sphere intersects the visual hull, it is removed from the following steps and it remains in that fixed reached position. The iteration is stopped when all points are fixed or are lost (e.g. when the center of mass is outside the object volume). The step is variable because, when a point is back-projected near the silhouettes contours, the step is reduced to reach a better approximation of the object model.

It is important to underline that the precision of reconstruction depends, obviously, on the number of views, on the observation angle and distance, and on the sphere sample points density. On the other hand, with the increase of N and d, the computational time of the algorithm increases too. A right trade-off may be easily found for each specific application.

Differently from the algorithm proposed in [11], the proposed method does not reconstruct the entire volume of the object, while only the object surface is estimated saving computational time. In this way, the use of voxels and their subdivision in other voxels, when one of these intersects the visual hull, becomes useless in this context. Moreover, the sample points are not labeled during the acquisition process as voxels in [11] but remain only fixed when they are back-projected in all silhouettes further simplifying the process environment.

IV. PROJECTION AND SHADOW CONE DISTURBANCES REJECTION

A new method to avoid disturbances due to projection and shadow cones, typical noises in these applications, is proposed in this section. This approach is based on an estimation of the object size along the revolution cone axis so to avoid that some points may be fixed outside these estimated limits of the object during the step-by-step iteration.

By denoting with $\mathbf{z}$ the revolution cone axis, the first step of this estimation method consists in projecting this axis onto all the image plane of all the acquired images. Therefore, in
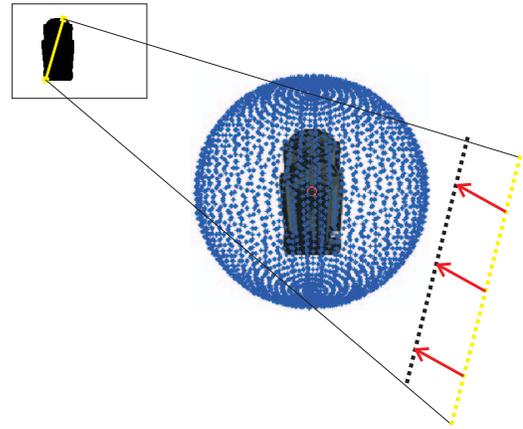

Fig. 5. The figure shows the direction of approach of the straight line towards the centroid and the image plane.
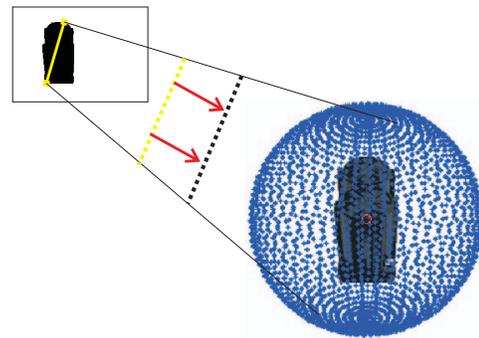

Fig. 6. The figure shows the direction of approach of the projected segment towards the centroid and fair from image plane.

each image plane, a straight line $\mathbf{z}_{pi}$ is found, corresponding to the projection of $\mathbf{z}$ on the plane i-th. The second step consists in finding the segment with the maximum length contained into the silhouette and parallel to the direction of $\mathbf{z}_{pi}$, as shown in Fig. 4.

The segment is sampled and its points $\mathbf{s}_{ij}$, with $j = 1, ..., h$ (h is the number of points used to sample the segment), are projected into the 3D space outside the estimated volume of the object and at the opposite part of the centroid with respect to the current camera (see Fig. 5). All these points are brought towards the centroid with a fixed step size η, as shown in Fig. 5. At each step, each point of the segment is back-projected onto all the image planes. Once the projection of a point is inside all the silhouettes, this point is fixed and is not considered for the next iterations. The process is stopped when all points have been fixed or do not belong to the object. Finally, the set of all the fixed points is processed to find the sample point $\mathbf{s}_{ij}$ corresponding to the maximum value along the positive direction of $\mathbf{z}$ that belongs to the object.

Repeating this process for all the N views, a vector of all the maximum points $\mathbf{M}_s$ can be composed. Qualitatively these values represent the coordinate of the higher point of the object seen from each view along the positive direction of the revolution cone axis $\mathbf{z}$. Since each view can be affected
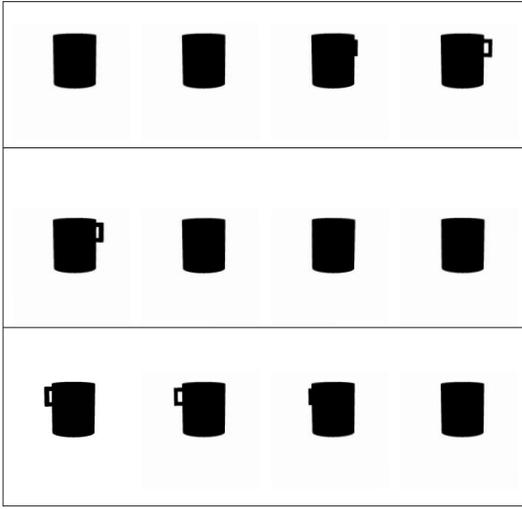
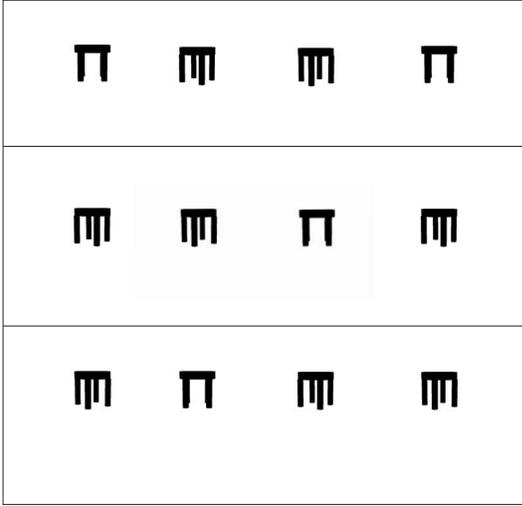Fig. 7. Collection of images generated for the solid cup with $\alpha = 80^\circ$.

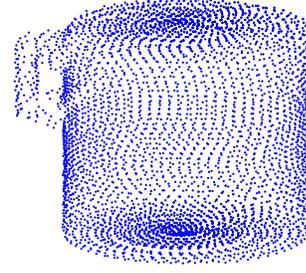Fig. 8. Collection of images generated for the table with $\alpha = 80^\circ$.

Fig. 9. Points reconstruction of the solid cup generated with MATLAB and reconstructed using 12 images acquired from a circular path surrounding the object with $\alpha = 80^\circ$.
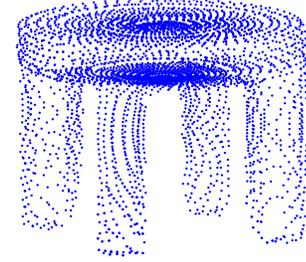

Fig. 10. Points reconstruction of the table generated with MATLAB and reconstructed using 12 images acquired from a circular path surrounding the object with $\alpha = 80^\circ$.

the images and corresponding to the maximum value along the negative direction of $\mathbf{z}$, and δ_m is a security coefficient depending on the step size η and used to avoid that a piece of the object is cut off due to image noise. The value $|M - m|$ is an estimation of the object height along the direction $\mathbf{z}$.

All fixed points considered in the object model reconstruction process that have a value along $\mathbf{z}$ bigger than M and lower than m can not be marked as fixed during the iterative reconstruction process and so the iteration process has to be continued.

V. SIMULATIONS AND EXPERIMENTS

A. Simulations

To test the algorithm with simulations, collections of images have been generated from object models suitably synthesized, and then used to reconstruct the 3D models with the proposed method.

A solid cup and a table have been synthesized in the MATLAB environment. $N = 12$ images are generated on a circular path with $\alpha = 80$ (see Fig. 1). For each point of view selected on the circular path an image plane is fixed, and in each of them the synthesized objects have been projected to obtain the silhouettes, representing the desired images.

In Fig. 7 and in Fig. 8 the set of the images taken from the considered circular path are shown. Notice that the silhouettes generated in his way are ideal, without problems of noise, illumination, etc.

The reconstruction sphere is made up of about 6.000 points equally distributed on it, while the other reconstruction

by noise and disturbance of projection and shadow cones, the value

$$M = \min \mathbf{M}_s + \delta_M$$

can be considered as the coordinate of the object highest point, with δ_M a weighting coefficient, related to the step size η, used to avoid that a piece of the object is cut off due to image noise.

A similar process can be employed to find the extension of the object along the negative direction of $\mathbf{z}$. The process is quite similar to the previous one. In this case, the sampled segment is projected outside the object volume but between the object and the camera, and it is moved away from the image plane towards the centroid, as shown in Fig. 6. In the same way as before, the coordinate of the object lowest point may be estimated as follows

$$m = \max \mathbf{m}_s + \delta_m,$$

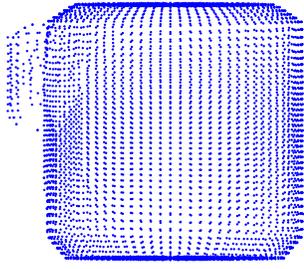where $\mathbf{m}_s$ is the vector of all the points extracted from all

Fig. 11. Points reconstruction of the solid cup generated with MATLAB and reconstructed using 12 images acquired from a circular path surrounding the object with $\alpha = 80°$, without using the disturbances rejection algorithm proposed in section IV.

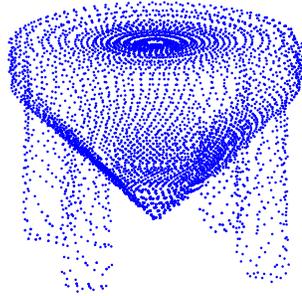Fig. 12. Effects of the projection cone on a non-convex object ($\alpha = 45°$).

parameters are as follows: $\epsilon = 0.01$, $\eta = 0.01$, and $\delta_m = \delta_M = 0.01\eta$.

In Fig. 9 and Fig. 10 the reconstructed models of the solid cup and of the table are shown; the goodness of the algorithm is evident due also to the quality of the available silhouettes. Fig. 11 illustrates the reconstruction results without the estimation of object dimensions proposed in section IV.

On the other hand, Figure 12 shows the crucial importance in the choice of α, especially when non-convex objects are considered. In this case a value $\alpha = 45°$ is chosen, so there is not the possibility to know what is present under the table and the corresponding projection cone is very conspicuous.

B. Experiments

A real case has been considered to fully appreciate the performance of the proposed algorithm.

The setup available in the PRISMA Lab employed for the proposed experiment consists of an industrial robot CO-MAU SMART-3S, with six-revolute-joint anthropomorphic geometry, a non-null shoulder and elbow offsets, and non-spherical wrist, and of a vision system, composed of a PC with Pentium IV 2.0GHz processor and a SONY 8500CE B/W camera, mounted in eye-in-hand configuration on the robot.

The target object is a supply battery, $N = 12$ is the number of desired views, and $\alpha = 45°$ is chosen to test and stress the proposed algorithm in an unfavorable condition that increases the effects of the projection cone. The circular acquisition trajectory is automatically generated and followed by the
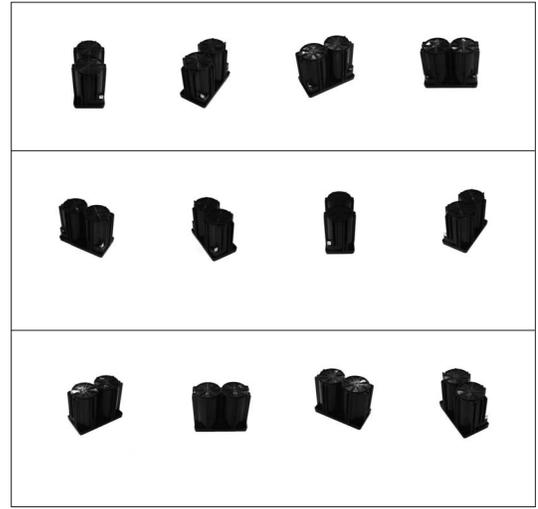

Fig. 13. Collection of the 12 images of the object taken by the robot with $\alpha = 45°$.
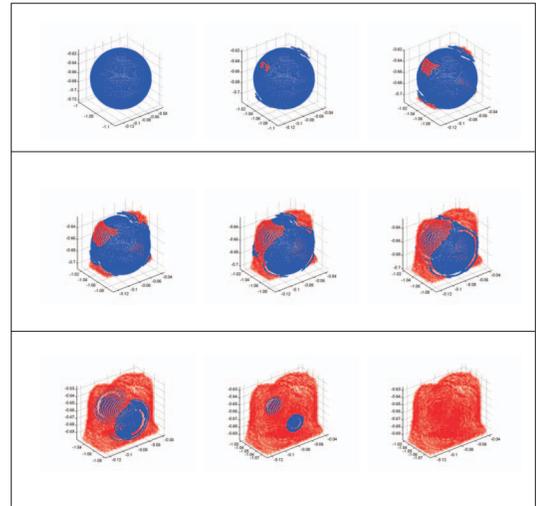

Fig. 14. Collection of images representing some of the steps of object reconstruction from initial condition to the final reconstruction.

robot with the camera mounted on the hand: the results of the acquired images are represented in Fig. 13. Obviously, with respect to the simulation case the image processing is more difficult due to disturbances like shadow, light differences and so on.

The initial reconstruction sphere is made up of about 10.000 points equally distributed, while the other reconstruction parameters are chosen equal to the simulation case.

The images sequence of the object model reconstruction process is shown in Fig. 14, while in Fig. 15 are represented the reconstruction results. The final object model is a good approximation of the real one and this is a further proof of the correctness of the algorithm even in a real situation with image noises and with an unfavorable choice of α, as said before.

Moreover, the reconstruction model without the use of the disturbance rejection algorithm is reported in Fig. 16 to show
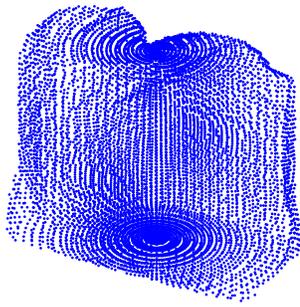
Fig. 15. Points reconstruction of a real object observed by a robot with an eye-in-hand camera configuration and reconstructed starting form 12 images taken from a circular path surrounding the object with an angle of $\alpha = 45^o$.
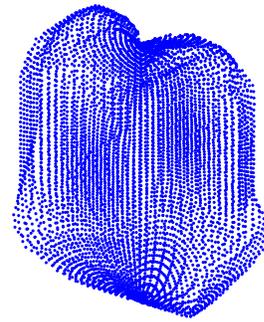

Fig. 16. Points reconstruction of a real object observed by a robot with an eye-in-hand camera configuration and reconstructed starting form 12 images taken from a circular path surrounding the object with an angle of $\alpha = 45^o$, without using the estimation of the object dimension.

the effectiveness of this solution also in a real case.

Finally, notice that this technique, based on silhouettes and more on blob analysis, does not allow a perfect reconstruction of an object model with concavities and grooves [6].

VI. CONCLUSION AND FUTURE WORK

A. Conclusion

In this paper a method for object model surface reconstruction has been proposed, which aims at acquiring information on the robot environment reducing the complexity of the algorithm with respect to others to allow a reconstruction process in real-time. This solution may be considered, for example, to perform real-time robotic grasp planning, where the time is an important parameter, more than the quality of the reconstruction.

The proposed method uses points to reconstruct the object surface and it employs a new algorithm to estimate the object dimensions so as to reduce the disturbance effects of shadow and projection cones.

Simulations and real experiments have been presented to demonstrate the effectiveness of the reconstruction process, from which it is clear that a trade-off between the number of views and the points density of the reconstruction sphere has to be taken into account, as well as the observation angle that could degrade object model reconstruction especially in the case of a non-convex object.

B. Future Work

Future developments of the algorithm are addressed especially towards the formulation of a real-time robotic grasping planning algorithm. Moreover, the enhancement that the addition of images taken from different values for the observation angle can give to the algorithm has to be further investigated, especially to refine and improve the dimension estimation method used to reject the reconstruction noise due to the projection cone and shadows.

VII. ACKNOWLEDGMENTS

REFERENCES

[1] R. Cipolla and A. Blake, "Surface shape from the deformation of apparent contours", *Internation Journal of Computer Vision*, vol. 9, no. 2, pp. 83-112, 1992.

[2] C. R. Dyer, "Volumetric scene reconstruction from multiple views", *Foundations of Image Analysis*, L.S. Davis ed., Kluwer, Boston, 2001.

[3] J.-S. Franco, E. Boyer "Exact polyhedral visual hulls", *British Machine Vision Conference*, 2003.

[4] T. H. Hong and M. O. Shneier, "Decribing a robot's workspace using a sequence of views from a moving camera", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 7, no. 6, pp. 721-726, 1985.

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

[6] A. Laurentini, "How far 3D shapes can be understood from 2D silhouettes", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 2, pp. 188–195, 1995.

[7] W. E. Lorensen and H. E. Cline, "Marching cubes: a high resolution 3D surface construction algorithm", *Computer Graphics*, vol. 21, no. 4, pp. 163–169, 1987.

[8] H. Noborio, S. Fukuda and S. Arimoto, "Construction of the octree approximating three-dimensional objects by using multiple views", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, no. 6, pp. 769–782, 1988.

[9] S. Prakoonwit and R. Benjamin, "3D surface point and wireframe reconstruction from multiview photographic images", *Image and Vision Computing*, vol. 25, pp. 1509–1518, 2007.

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

[11] R. Szeliski, "Rapid octree construction from image sequences", *CVGIP: Image Understanding*, vol. 58, no. 1, pp. 23–32, 1993.

[12] K. K. Wong and R. Cipolla, "Reconstruction of outdoor sculptures from silhouettes under approximate circular motion of an uncalibrated hand-held camera", *IEICE Transactions on Information and Systems*, vol. E87-D, no. 1, 2004.

[13] C. Xu and J. L. Prince, "Snakes, shapes, and gradient vector flow", *IEEE Transactions on Image Processing*, vol. 7, no. 3, 1998.

[14] A. Yezzi, G. Slabaugh, A. Broadhurst, R. Cipolla and R. Schafer, "A surface evolution approach to probabilistic space carving", *Proceedings of the First International Symposium on 3D Data Processing Visualization and Trasmission*, 2002.

[15] T. Yoshikawa, M. Koeda and H. Fujimoto, "Shape recognition and optimal grasping of unknown objects by soft-fingered robotic hands with camera", *11th International Symposium on Experimental Robotics*, Athens, 2008.

[16] B. H. Yoshimi and P. K. Allen, "Visual control of grasping and manipulation tasks", *Proceedings of the 1994 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, Las Vegas, 1994.