PROBABILISTIC SILHOUETTE-BASED CLOSE-RANGE PHOTOGRAMMETRY USING A NOVEL 3D OCCUPANCY-BASED RECONSTRUCTION

I. Asl Sabbaghian Hokmabadi*, N. El-Sheimy

Dept. of Geomatics Engineering, University of Calgary, Calgary, Canada
(ilyar.aslsabbaghianh, elsheimy@ucalgary.ca)

KEY WORDS: Silhouette-based 3D Reconstruction, Probabilistic 3D Reconstruction, Dense Reconstruction, 3D Occupancy Grids, Virtual Bounding Box, Fisher's Linear Discriminant Function.

ABSTRACT:

Digital three-dimensional (3D) reconstruction of objects has many applications in computer vision, archaeology, and the entertainment industry. Digital 3D reconstruction can be used to preserve the appearance of valuable historical artifacts; it can be used to track the pose of an object in the images, and it can facilitate object modelling. 3D reconstruction of objects in the past has been achieved using many sensors such as cameras and laser-strip scanners. Monocular camera-based object 3D modelling can be categorized into sparse feature detector/descriptor-based and dense silhouette-based approaches. Feature-based methods identify distinctive features on the objects (captured from many images). In contrast, silhouette-based methods only require a distinguishable boundary between the object and the background. Silhouette-based methods have the advantage that in the controlled setups, a special background can be designed to be distinguishable from the object of interest; therefore, uniquely identifiable textures on the object's surface are not required. Despite their advantages, silhouette-based probabilistic reconstruction remains a challenge. This article proposes a new probabilistic approach using 3D occupancy grids for the silhouette-based digital reconstruction of an object. The proposed method is designed to be usable with monocular cameras and achieves an accurate reconstruction using only sixteen images. Compared to similar silhouette-based volumetric approaches, the voxels are not discarded immediately during the reconstruction, and the occupancy grid mapping continuously changes the occupancy probability of the voxels with each new image included.

1. INTRODUCTION

Object 3D digital reconstruction has been a topic of interest for many years. 3D reconstruction can preserve valuable historical and archaeological artifacts in digital format (Van Nguyen et al., 2021). In the entertainment industry, it can accelerate 3D modelling (Statham, 2020). More recently, it has been utilized to track targets using images (Majcher and Kwolek, 2020).

Objects in the past have been reconstructed digitally using different types of sensors and setups (e.g., stereo-pair, sensor array) such as laser-strip scanners (Curless and Levoy, 1996), monocular cameras (Pollefeys et al., 1999), RGB-D cameras (Anasosalu, 2013), multiple stereo-pairs (Peng et al., 2015), and structured light (Ullah et al., 2020). Among the different types of sensors, cameras can preserve the appearance of the object (e.g., colour, texture). Further, using a monocular camera compared to multi-camera and stereo setups can reduce the cost of the equipment if the accuracy of the reconstruction can meet the requirements for the application.

Digital reconstruction methods using a monocular camera can be categorized into feature point detector/descriptor-based (Fang et al., 2020) and silhouette-based methods (Bandyonadhyay et al., 2019).

Feature point detector/descriptors identify and correspond distinctive points on the object. These special feature points can be external markers (Hou, 2022) or uniquely identifiable textures on the object's surface (Ummenhofer and Brox, 2013). Feature detection/descriptor algorithms such as Scale Invariant Feature Transformation (SIFT) (Lowe, 2004) and Speeded-up Robust Features (SURF) (Bay at el., 2006) are amongst the methods used for automated feature point detection. In contrast to detecting features on the objects, silhouette-based reconstruction depends on a distinguishable boundary between the object (foreground) and the background.

Silhouette-based 3D reconstruction has several advantages over feature detector/descriptor methods. Among these advantages is that silhouette-based reconstruction can be used for objects that do not naturally have a textured surface but can be distinguished from their background (e.g., using an object's colour). In controlled setups (such is the case with this paper), the background can be specifically designed to be distinctive from the object (foreground). The second advantage of silhouette-based 3D reconstruction is that it provides a denser representation of the object of interest compared to the feature detector/descriptors approach.

Silhouette-based 3D reconstruction can be used to represent an object using several different modelling techniques. For example, objects can be represented using explicit parameterization (Esteban and Schmitt, 2004), B-splines (Wong and Cipolla, 2001), level-sets (Whitaker, 1998), and voxel-based volumes (Potmesil, 1987). The method proposed in this paper uses the voxel-based volumetric representation of the object, which has the advantage that it can represent arbitrary solid objects easier (Tosovic, 2002).

Despite the advantages, silhouette-based reconstruction has its drawbacks. One such drawback is that silhouette-based 3D reconstruction is limited to the visual hull of an object (Laurentini, 1994). The visual hull might not capture the concavities on the surface of an object (Lee and Yilmaz, 2010). The second difficulty is a probabilistic representation of the silhouettes. In most past literature work, voxels were removed or kept during the reconstruction of the silhouette, and the decision to remove pixels is deterministic (Mulayim et al., 2003, Kim et al., 2007, Gouiaa and Meunier, 2014, Bandyonadhyay et al., 2019). These approaches can be

* Corresponding author

https://doi.org/10.5194/isprs-archives-XLIII-B2-2022-343-2022 |

successful if the objects can be accurately segmented from the background in each image. Such accuracy might not be possible in every condition. For example, room illumination conditions can pose difficulties in detecting the entire foreground area in one or several consecutive images, leading to permanently discarding object voxels. The alternative approach to the deterministic removal of the pixels is to build a probabilistic voxel-based representation of the object (De Bonet and Viola, 1999, Bhotika et al., 2002, Franco and Boyer, 2005, Vogiatzis and Hernández, 2011, Kolev et al., 2014). While these methods can improve the deterministic approach, most rely on global (non-iterative) optimization to fit the best 3D model to all the images. Therefore, adding new images to improve the model's accuracy would require solving a large optimization problem again. Also, the methods in the past were designed for stereo-pair or multiple cameras.

The proposed method of this paper avoids removing any voxels during the reconstruction by formulating voxel-based volumetric silhouette modelling as a well-known mapping technique of 3D occupancy grids (Elfes, 1989). In the occupancy mapping, the voxels are all assigned to an occupancy value that can vary with the introduction of each new image. This iterative process creates a possibility of detecting missed foreground pixels in several views (but still detected in some other images). To the best of the author's knowledge, occupancy grids for accurate 3D reconstruction of the object using silhouettes with a monocular camera have not been proposed in the past.

Compared to the previous silhouette-based methods, projecting the voxels onto the image is avoided in the proposed approach. Such projection should be applied to all the voxels inside the bounding volume; since it is unknown which voxels belong to foreground/background segments. Conversely, in this paper, the pixels are back-projected to rays, intersecting a virtual bounding box designated around the object. The intersected segment of the ray is divided into equally spaced points. Finally, these points are assigned to the closest voxels. Since foreground pixels are detected in the image, it is possible to only back project the corresponding rays. The remaining voxels can be assumed to correspond to background pixels without intersection with any rays.

For the purpose of the experiment, it is assumed that the orientation of the monocular with respect to the object of interest is known in each new frame. This assumption is reasonable when a turntable with an accurate rotation angle is utilized to capture the images from the object of interest.

The images of the experiments are captured using Raspberry Pi Camera Module 2, which is a very low-cost sensor compared to high-end metric cameras. The number of images captured from this sensor for each reconstruction is sixteen.

2. METHODOLOGY

2.1 Overview

The proposed methodology relies on four major steps. Initially, the pose of the monocular camera with respect to the rotating platform is calculated using the first image. In the second step, a background/foreground segmentation is built. These two steps are performed only once as an offline stage. The third step involves intersecting the back-projected rays passing through the camera centre with the virtual box. Finally, in the fourth step, the occupancy values of the voxels are updated. These last two steps are performed iteratively for each new image captured from the object.

2.2 Camera Pose Estimation

In the first step, the pose of the monocular camera with respect to the turntable is calculated. In order to find this pose, the homography between the first image and the turntable's plane is estimated. A four-point algorithm is used to derive this homography (Hartley and Zisserman, 2009). See Figure 1 for the illustration, where the virtual grid points (blue points) on the turntable's plane are projected (using the homography) onto the first image. The red and green colours denote the x and y axes in the world frame (the black point corresponds to the frame's centre). These axes are attached to the object (they rotate as the object rotates), and the object is reconstructed in this frame.

The estimated homography can be decomposed to find the translation and the rotation matrix of the monocular in the first view. Assuming that the intrinsic calibration parameters of the camera are known, the method in (Zhang, 2000) can be used for the decomposition. The camera's subsequent relative poses (to the object's frame) are calculated using the known accurate angle of rotation of the turntable. Figure 2 illustrates this for sixteen camera frames (red, blue, and green colours denote the x, z and y axes). The axes are scaled such that the distance from the centre of the platform to each corner represents 1 unit. The first view in Figure 2 corresponds to the image in Figure 1. The blue points are in the x-y plane of the world frame, corresponding to the blue points in Figure 1(the number of points is reduced in Figure 2 for clear illustration).

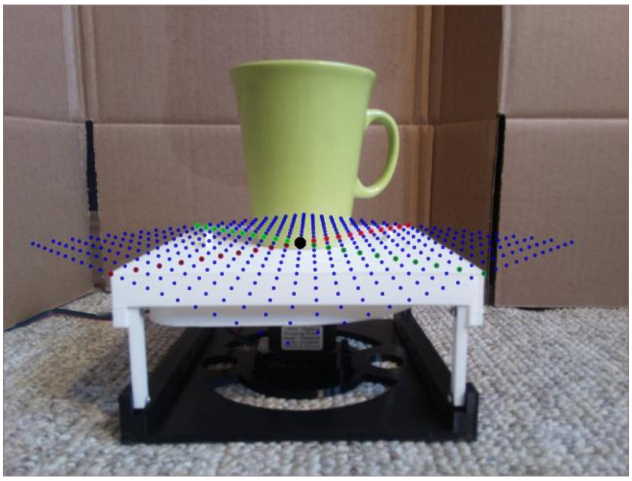

Figure 1. Illustration of homography-based projection. Blue points are the projection of the world coordinates in the x-y plane. Red and blue points are the x and y axes in the world frame. The black point is the centre of the world frame.

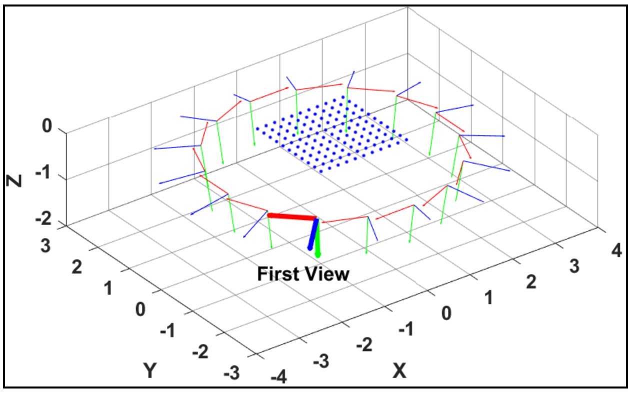

Figure 2. Sixteen camera poses are shown in the world frame. First-view corresponds to the image in Figure 1.

2.3 Foreground and Background Segmentation

Fisher's linear discriminant function (Fisher, 1936) is used to segment the background and the foreground. This approach can be considered a supervised classification method. The training for this function required as few as twelve points (each for foreground and background) in the first image. Once the training is achieved, the resultant discriminant function can be applied to the subsequent images.

Figure 3 demonstrates applying the discriminant function to the first and seventh images in one of the datasets collected for the experiments. Black and white colours correspond to the background/foreground segments, respectively. The segmentation is mostly successful in both frames, even though the training is only achieved using very few points (and in the first image). Several falsely detected foreground pixels can be seen in Figure 3 (see white pixels at the lower left). However, the reconstruction area is limited to a virtual bounding box around the object in the experiments. Therefore, false positives such as these pixels are not included in the reconstruction. Besides false positives, there also exist false negatives (see black pixels on the cup's handle). One reason for the missing foreground pixels can be due to the illumination conditions. The errors due to the illumination might not be persistent in the same points on the object as the platform rotates. The results of 3D reconstruction in section 3 confirm this.

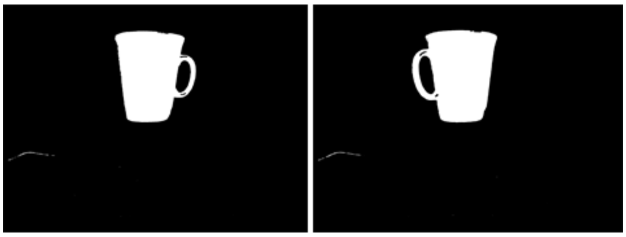

Figure 3. Examples of the binary images of the background and foreground segments

2.4 Virtual Box Intersection

In the third step, image pixels are back-projected to the scene and intersected with a virtual bounding box defined around the object. Back-projected rays that pass through the virtual box are assigned to the closest voxels. In the following, a simple algorithm to achieve this is explained.

Equation 1 shows a point r on the back-projected ray where x is the homogenous image pixel coordinates; P^{+} is the pseudo inverse of the camera matrix, and C is the camera's perspective centre in the world coordinate frame. Using Plücker's line representation, a line (L) passing through the perspective centre and the point r on the ray is shown in (2) (where $C^{'}$ is the transpose of C). To find the intersection of this line with the planes (M_i) of the virtual box, (3) can be utilized (Hartley and Zisserman, 2009). If a line passes through the virtual box, it must intersect this box at two special points; denoted as "entry" and "exit" points. Intersected points can be tested using the two conditions shown in (4) and (5), where U^{out} and U^{in} are the sets of the points outside and inside the virtual box, respectively. X^{+} and X^{-} are defined in Equations 6 (a and b), where $d(L)$ denotes the direction of the line and η is a small value.

$$r(\lambda) = P^{+}x + \lambda C \tag{1}$$

$$L = rC^{'} - Cr^{'} \tag{2}$$

$$X = LM_i, M_i \in \{1...6\} \tag{3}$$

$$(X^{+} \in U^{out} \wedge X^{-} \in U^{in}) \rightarrow X \text{ is an exit point} \tag{4}$$

$$(X^{-} \in U^{out} \wedge X^{+} \in U^{in}) \rightarrow X \text{ is an entry point} \tag{5}$$

$$X^{+} = X + \eta d(L), X^{-} = X - \eta d(L) \tag{6.a, 6.b}$$

2.5 Occupancy Grid Update

The estimated entry and exit points can be used to select points on each ray that is inside the bounding box. These points are then assigned to the closest voxels. In the last step of the proposed methods, the algorithm uses a formalization of the occupancy grid approach (Choset et al., 2005) to update the value of each voxel. The update equation is shown in (7), where the occupancy probability is denoted as p, and the logit function is denoted as l. The random variables representing the map, observations and the camera's pose are denoted as m, z, and s respectively. The observations are the pixels with binary values (see Figure 3). The superscript index k denotes the entire sequence starting from the initial state. The subscript index denotes only the last state.

The term on the left-hand side in (7) is the logit of the posterior of the occupancy grid. The terms on the right-hand are (in order), logit of the update, prior, and initial map probabilities. The initial term is 0, which corresponds to assigning an occupancy probability of 0.5 to each voxel. The prior term of the step k is the same as the posterior of $k-1$. The last term (update) will be different for the foreground and the background pixels shown in (8) (where F and B denote the set of foreground and background pixels). Based on Equation 8, the associated map voxels will be assigned to a higher occupancy value (than 0.5) for a foreground pixel. Conversely, voxels associated with background pixels will be assigned to a lower occupancy value.

$$l(p(m \mid z^{k}, s^{k})) = l(p(m \mid z_k, s_k)) + l(p(m \mid z^{k-1}, s^{k-1})) - l(p(m)) \tag{7}$$

$$\begin{cases} z_k \in F \rightarrow p(m \mid z_k, s_k) = 0.55 \\ z_k \in B \rightarrow p(m \mid z_k, s_k) = 0.45 \end{cases} \tag{8}$$

The four steps of the proposed algorithm are summarized in Figure 4. Camera pose detection and building background/foreground detector are performed as offline steps. In the online step, occupancy grids are built iteratively. In addition, a threshold can be used to designate the voxels with a higher probability as the 3D reconstruction of the object.

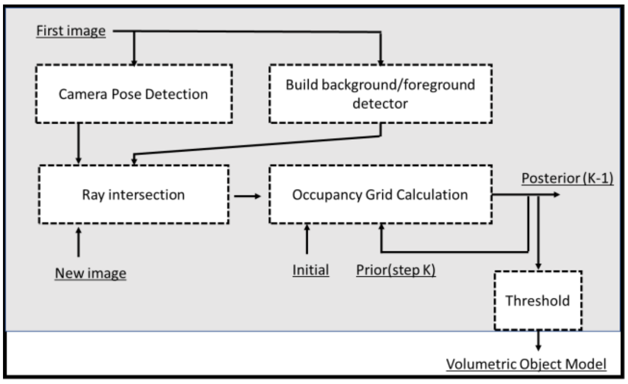

Figure 4. Flowchart of the proposed method

3. RESULTS

We have tested the accuracy of the proposed method using three image sets from different objects. For the first dataset, sixteen images are captured using a monocular camera. Figure 5 shows examples of the first set (four images are shown of the total sixteen images).

The calibration parameters of the camera are derived using a planar checkerboard and ©MATLAB's calibration toolbox. The relative translations and the rotations between the monocular camera and the turntable are calculated using the homography decomposition method (see section 2.2). The foreground pixels are detected (see section 2.3), and corresponded rays are intersected with a virtual box (see section 2.4). One in every 2 pixels is back-projected. However, for higher accuracy, all the pixels can be included, which would increase the memory requirements.

The dimensions of the virtual box for the first experiment are $261 \times 241 \times 217$ (in x, y and z axes). These dimensions are varied slightly in each direction to accommodate the size variation of the object in the second and third experiments.

An example of the intersected virtual box around the object for one image is shown in Figure 6, where occupied and unoccupied cells are in blue and red colours (the scale of this figure is unit world frame). The foreground rays are back-projected to more than one point. These include all the points along the path from the entry to exit points.

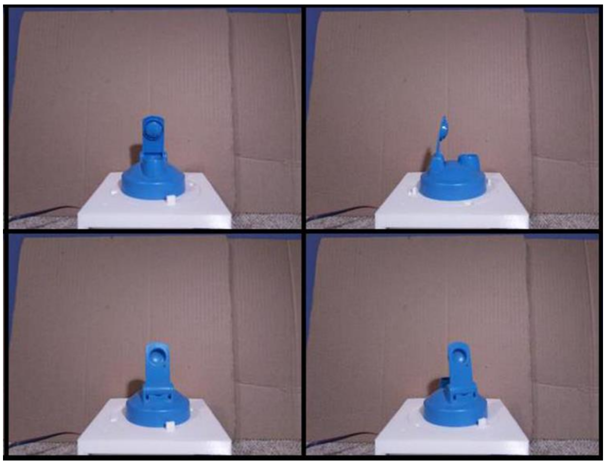

Figure 5. Example images from the first experiment

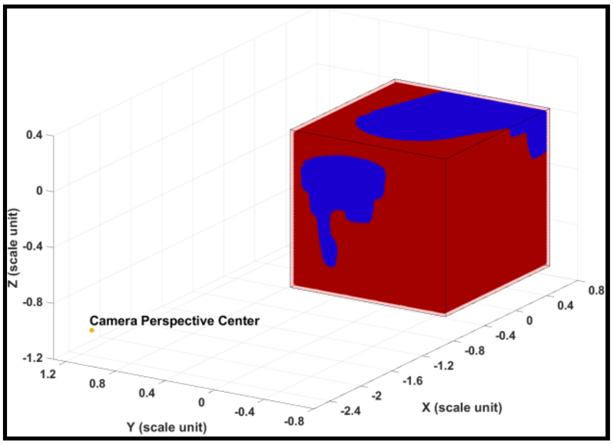

Figure 6. Back-projected object voxels in the virtual box. Each ray intersects the virtual box at many points. See the entry voxels in the front and some of the exit points on the top and side of the virtual box.

The occupancy grid voxels have a value between zero to one, as they are probabilities. A threshold can be set to only keep certain voxels with a higher probability value. This threshold is found experimentally in this article, but its value does not vary significantly between the experiments. For the first example, the threshold is set to 0.92. The reconstructed 3D model for the first set is shown in two views in Figures 7 and 8. The accuracy of the reconstructed model is compared to measurements obtained from the object using a calliper (with 0.02mm accuracy). The first measurement taken is the average radius of the object's base as 9.7 centimetres (cm). The second measurement is taken from the top to the base height of the object, and the value reported is 10.9 cm. The corresponding distances are calculated in the reconstructed model as 9.9 cm and 10.6 cm, respectively.

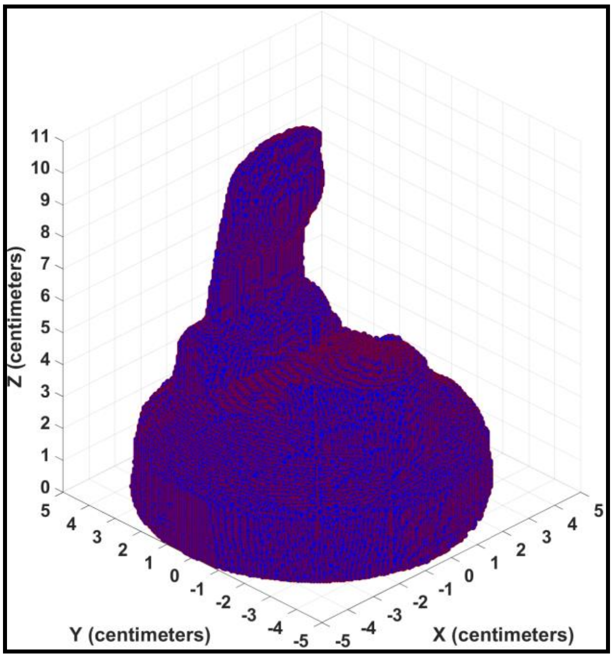

Figure 7. Reconstructed model (view 1)

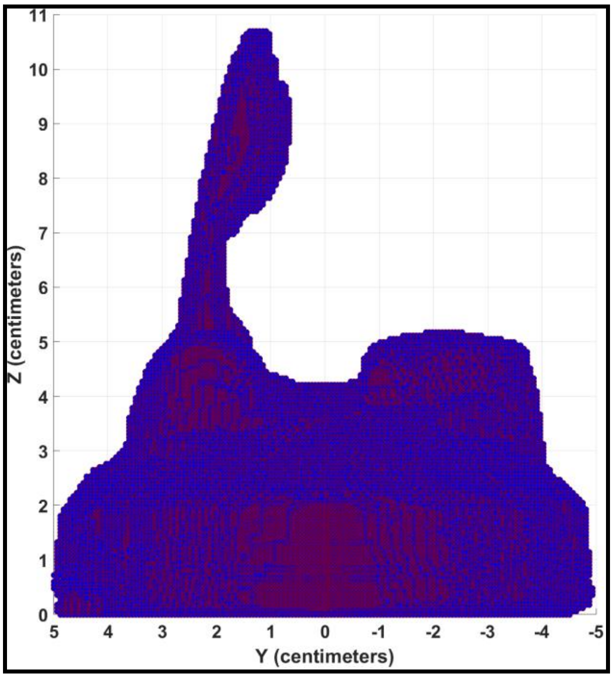

Figure 8. Reconstructed model (view 2)

As it was the main objective of this paper, the proposed method provides a complete probabilistic silhouette map in the reconstruction process. Figure 9 (b) shows the 3D probability map projected onto the z-y plane (The image of the object is shown in Figure 9 (a) as the reference). Since many voxels projects to one point, the maximum occupancy probability is used for the depictions. The thresholds in Figures 9 (c-f) are set to 0.96, 0.92, 0.91, and 0.8. Lowering the probability threshold will include more pixels with lesser occupancy probability, and the object looks to be scaled (see Figure 9 (f)). The change in the threshold has more effect in the y-direction compared to the z-direction (the object looks more scaled in the y-direction). A possible solution to increase the difference between the occupancy values of the object and non-object voxels is to use more images from new viewpoints. For this experiment, if an image is acquired with the camera placed above the object, voxels that fall beyond the object's base will receive a lower occupancy probability.

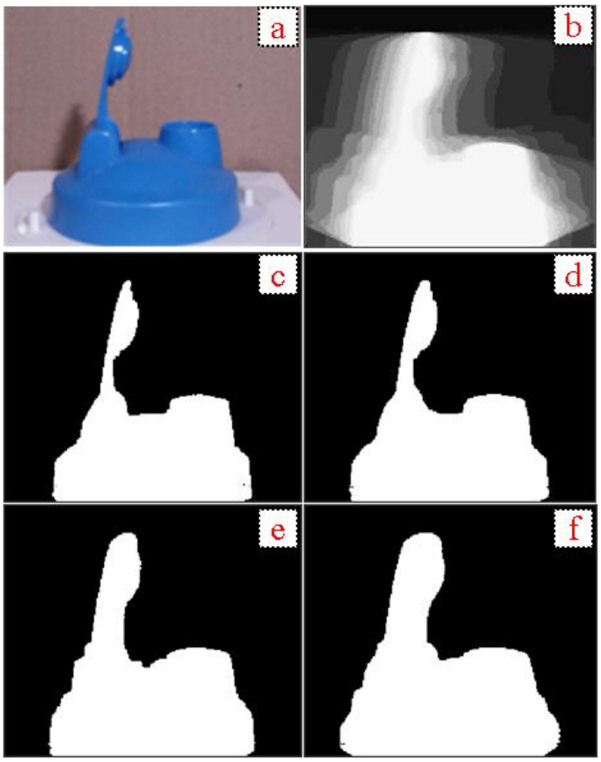

Figure 9. Illustration of the effect of varying thresholds on the object reconstruction. Figure 9 (a) shows the image of the object. Figure 9 (b) is the projected probability map onto the z-y plane. Figures 9 (c-f) correspond to the threshold values of 0.96, 0.92, 0.91 and 0.8.

Example images from the second set are shown in Figure 10. To further explore the effect of the varying threshold on the model, reconstruction using three threshold values is shown in Figure 11 (a, b, and c correspond to the threshold values of 0.91, 0.94 and 0.96). The area of the most interest for the discussion is highlighted inside the red box. The lower threshold includes the missed voxels on the cup's handle (see Figures 11 (a) and 11 (b)). However, for the threshold value of 0.96, some of the voxels on the cup's handle are missing due to segmentation errors. A possible source of the error can be the high surface reflection on the cup's handle in some of the images.

Further reconstruction errors can be seen as the additional voxels on the top of all three models, which are not present on the object (Figure 11 (d)). The reason for this error is that the images captured do not observe the top section of the object (see images in Figure 10). Figure 12 shows a final model of the object where these voxels are removed manually.

The accuracy of the reconstruction is measured using several distances collected from the object and the reconstructed model. The reconstruction error is defined as the average of the absolute difference between the ground truth and the reconstruction measurements. These errors are reported as 0.15 cm, 0.4cm and 1.2 cm for 0.91, 0.94, and 0.96 thresholds, respectively.

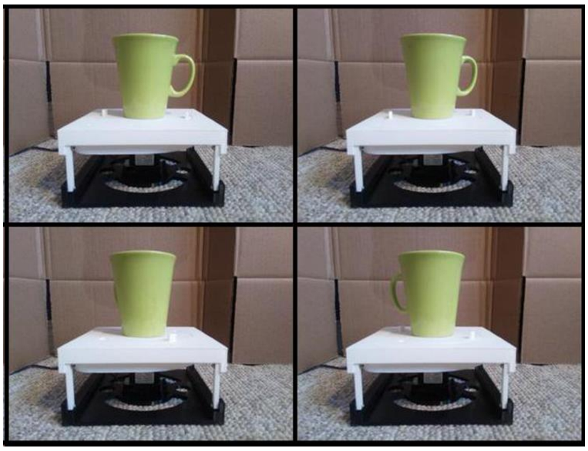

Figure 10. Example images from the second experiment

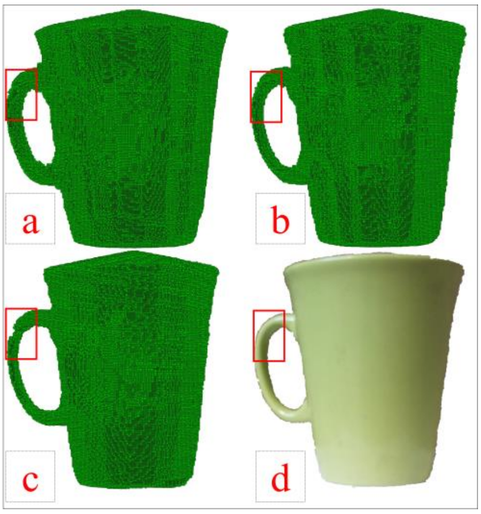

Figure 11. A comparison of the reconstruction accuracy for different thresholds. Figures 11 (a), 11 (b), 11 (c) correspond to the threshold values of 0.91, 0.94, and 0.96. Figure 11 (d) shows the image of the object as the reference.

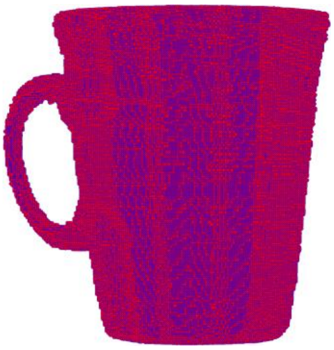

Figure 12. Final reconstructed model (second experiment)

The last experiment is performed using the object shown in Figure 13. A utility of the proposed probability reconstruction is the ability to incorporate the knowledge that can improve the accuracy of the final reconstruction. Such knowledge can be regarding the illumination of the room and/or a prior initial model of the object. In this experiment, the turntable includes extremities (highlighted inside the box in the image in Figure 14). These extremities block parts of the foreground at certain angles and reduce the occupancy values of the corresponding voxels. Since the occupancy grid method is iterative, it is possible to combine two or more already reconstructed models without any requirement to recalculate the original occupancy grids. While the turntable issue can be avoided with professional and expensive equipment, the iterative process of the occupancy grid can also provide a solution to accommodate other possible sources of error (e.g. errors due to illumination).

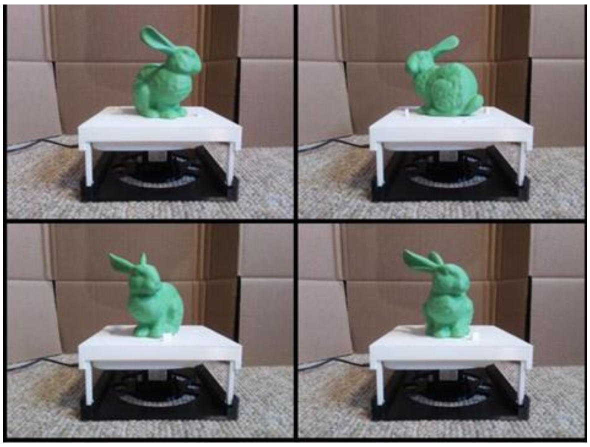

Figure 13. Example images from the third experiment

In Figure 14, two reconstructions are compared to each other. The reconstruction in Figure 14 (a) is built with the initial sixteen images. The model lacks voxels in the highlighted box region. A secondary reconstruction is built using another sixteen images captured at a higher camera angle where the extremity is not blocking the view. This secondary model is registered to the initial model using an Iterative Closest Point algorithm (Chen and Medioni, 1992). The probability values of the second model are utilized as the update term in Equation 7. The prior is the occupancy probability map of the initial model. The result is shown in Figure 14 (b), and it can be seen that the model has improved the accuracy of the initial reconstruction in the box.

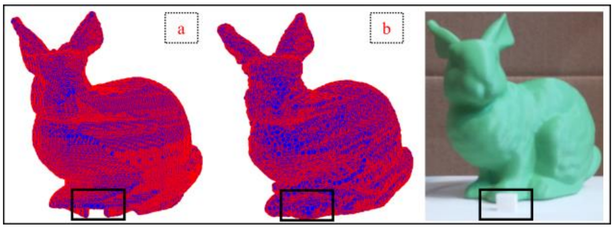

Figure 14. Comparison of the two reconstructed models. Figures (a and b) correspond to the initial and improved models. The area inside the box in the image shows the turntable extremity, which is a source of error.

The accuracy of the registration is very important in the quality of the final combined reconstruction. Unfortunately, the errors in the registration introduce distortions in the final model. The two views selected as samples of the final reconstruction in Figure 15 correspond to the secondary reconstruction alone.

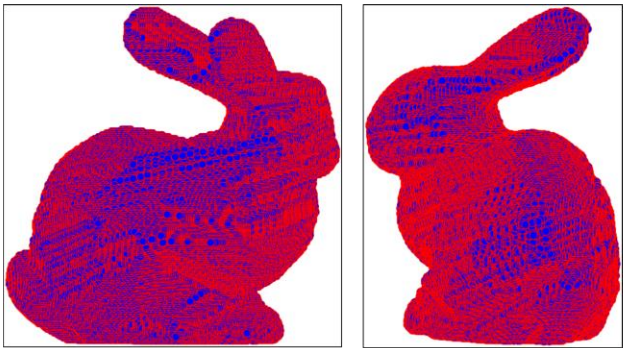

Figure 15. Two views of the reconstructed model (third experiment)

The reconstruction model has been compared to the Computer-Aided Design (CAD) in Figure 16. This CAD model was used to 3D print the object, and therefore it is an accurate representation of the object. Figure 16 (a) shows the selected view of the reconstruction. The mesh model of the reconstruction, the ground-truth model, and the superimposed models are shown in Figures 16 (b), 16 (c), and 16 (d), respectively.

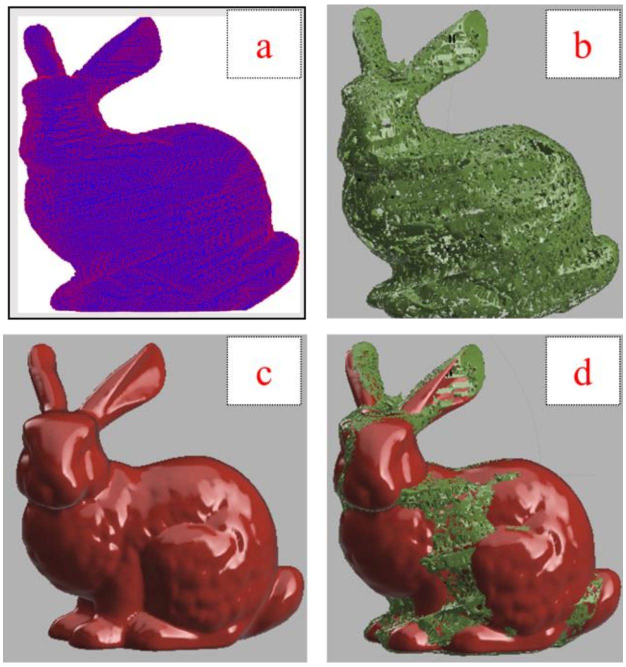

Figure 16. Comparison of the reconstruction and ground-truth model. Figures (a-d) show the reconstructed voxel-based modelled, mesh model, ground-truth model, and superimposed models, respectively.

4. CONCLUSIONS AND FUTURE WORK

In this paper, a new method for probabilistic silhouette-based 3D reconstruction is proposed. The proposed method utilizes the known mapping method of occupancy grids. Due to the iterative process of the occupancy grid-based reconstruction, new images can be incorporated into the reconstruction without the requirement to process the previous images again.

Most methods of silhouette-based reconstruction methods have utilized volumetric voxel-based representation. These approaches project the voxels in the bounding box onto the image plane. Instead, in this paper, it is proposed to back project only the foreground pixels, which helps detect the voxels with higher occupancy probability. The remainder of the voxels is assigned to lower occupancy probability without further computations.

The estimated accuracy of the proposed method is in the range of a few millimetres for the first two experiments. In the third experiment, the reconstruction model is compared to the CAD model of the object, where it exhibits an accurate representation of the object. This accuracy was achieved using a low-cost monocular camera and sixteen images per experiment.

The utility of the probabilistic reconstruction method has been discussed in this paper. The proposed probabilistic reconstruction provides a complete occupancy posterior of the bounding volume box. This posterior can be visualized to help find the weaknesses in the reconstruction. For example, if the objective is to maximize the difference between the occupancy value of the object and non-object voxels, the posterior can provide insights into where to localize the camera for new images. Further, it is demonstrated more than one reconstruction can be combined to mitigate certain errors. The requirements for this combination are registration between the reconstructed models and one occupancy grid update step.

In this article, the update term is assigned to a fixed value. This approach is based on the assumption that each observation can be assigned to either background or foreground deterministically. However, it is possible that different parts of the object's surface have a different level of saliency from the background. A more robust approach is to incorporate the uncertainty in the segmentation in assigning probability values for the voxels.

5. ACKNOWLEDGMENT

This research has been supported by funding of Prof. Naser El-Sheimy from NSERC CREATE and Canada Research Chairs programs.

REFERENCES

Anasosalu, Pavan, Diego Thomas, and Akihiro Sugimoto. "Compact and accurate 3-D face modeling using an RGB-D camera: let's open the door to 3-D video conference." In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 67-74. 2013.

Bandyonadhyay, Santarshi, Issa Nesnas, Shvam Bhaskaran, Beniamin Hockman, and Benjamin Morrell. "Silhouette-based 3d shape reconstruction of a small body from a spacecraft." In *2019 IEEE Aerospace Conference*, pp. 1-13. IEEE, 2019

Bay, Herbert, Tinne Tuytelaars, and Luc Van Gool. "Surf: Speeded up robust features." In *European conference on computer vision*, pp. 404-417. Springer, Berlin, Heidelberg, 2006.

Bhotika, Rahul, David J. Fleet, and Kiriakos N. Kutulakos. "A probabilistic theory of occupancy and emptiness." In *European conference on computer vision*, pp. 112-130. Springer, Berlin, Heidelberg, 2002.

Chen, Yang, and Gérard Medioni. "Object modelling by registration of multiple range images." *Image and vision computing* 10, no. 3 (1992): 145-155.

Choset, Howie, Kevin M. Lynch, Seth Hutchinson, George A. Kantor, and Wolfram Burgard. *Principles of robot motion: theory, algorithms, and implementations*. MIT press, 2005.

Curless, Brian, and Marc Levoy. "A volumetric method for building complex models from range images." In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pp. 303-312. 1996.

De Bonet, Jeremy S., and Paul Viola. "Poxels: Probabilistic voxelized volume reconstruction." In *Proceedings of International Conference on Computer Vision (ICCV)*, pp. 418-425. 1999.

Elfes, Alberto. "Using occupancy grids for mobile robot perception and navigation." *Computer* 22, no. 6 (1989): 46-57.
Esteban, Carlos Hernández, and Francis Schmitt. "Silhouette and stereo fusion for 3D object modeling." *Computer Vision and Image Understanding* 96, no. 3 (2004): 367-392.

Fang, Zhihao, He Ma, Xuemin Zhu, Xutao Guo, and Ruixin Zhou. "Sefm: A sequential feature point matching algorithm for object 3d reconstruction." *In International Conference on Frontier Computing*, pp. 283-296. Springer, Singapore, 2020.

Fisher, Ronald A. "The use of multiple measurements in taxonomic problems." *Annals of eugenics* 7, no. 2 (1936): 179-188.

Franco, J-S., and Edmond Boyer. "Fusion of multiview silhouette cues using a space occupancy grid." In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, vol. 2, pp. 1747-1753. IEEE, 2005.

Gouiaa, Rafik, and Jean Meunier. "3D reconstruction by fusioning shadow and silhouette information." In *2014 Canadian Conference on Computer and Robot Vision*, pp. 378-384. IEEE, 2014.

Hartley, Richard I. and Andrew Zisserman. "Multiple View Geometry." *Encyclopedia of Biometrics* (2009).

Hou, Guanyu, Weibin Zhang, Bin Wu, and Rongfang He. "3D reconstruction and positioning of surface features based on a monocular camera and geometric constraints." *Applied Optics* 61, no. 6 (2022): C27-C36.

Kim, Hansung, Ryuuki Sakamoto, Itaru Kitahara, Neal Orman, Tomoji Toriyama, and Kiyoshi Kogure. "Compensated visual hull for defective segmentation and occlusion." In *17th International Conference on Artificial Reality and Telexistence (ICAT 2007)*, pp. 210-217. IEEE, 2007.

Kolev, Kalin, Petri Tanskanen, Pablo Speciale, and Marc Pollefeys. "Turning mobile phones into 3D scanners." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3946-3953. 2014.

Laurentini, Aldo. "The visual hull concept for silhouette-based image understanding." *IEEE Transactions on pattern analysis and machine intelligence* 16, no. 2 (1994): 150-162.

Lee, Heewon, and Alper Yilmaz. "3D Reconstruction using Photo Consistency from Uncalibrated Multiple Views." In *VISAPP (1)*, pp. 484-487. 2010.

Lowe, David G. "Distinctive image features from scale-invariant keypoints." *International journal of computer vision* 60, no. 2 (2004): 91-110.

Majcher, Mateusz, and Bogdan Kwolek. “3D Model-based 6D Object Pose Tracking on RGB Images using Particle Filtering and Heuristic Optimization.” In *VISIGRAPP (5: VISAPP)*, pp. 690-697. 2020.

Mulayim, Adem Yasar, Ulas Yilmaz, and Volkan Atalay. “Silhouette-based 3-D model reconstruction from multiple images.” *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 33, no. 4 (2003): 582-591.

Peng, Qi, Lifen Tu, Kaibing Zhang, and Sidong Zhong. “Automated 3D scenes reconstruction using multiple stereo pairs from portable four-camera photographic measurement system.” *International Journal of Optics* 2015 (2015).

Pollefeys, Marc, Reinhard Koch, and Luc Van Gool. “Self-calibration and metric reconstruction inspite of varying and unknown intrinsic camera parameters.*” International Journal of Computer Vision 32, no. 1* (1999): 7-25.

Potmesil, Michael. “Generating octree models of 3D objects from their silhouettes in a sequence of images.” *Computer Vision, Graphics, and Image Processing 40,* no. 1 (1987): 1-29.

Statham, Nataska. “Use of photogrammetry in video games: a historical overview.” *Games and Culture* 15, no. 3 (2020): 289-307.

Tosovic, Srdan, Robert Sablatnig, and Martin Kampel. “On Combining Shape from Silhouette and Shape from Structured Light.” In *Proc. of 7th Computer Vision Winter Workshop, Bad Aussee*, pp. 108-118. 2002.

Ullah, Furqan, Sajjad Miran, Furqan Ahmad, and Irfan Ullah. “A low-cost three-dimensional reconstruction and monitoring system using digital fringe projection.” *Transactions of the Canadian Society for Mechanical Engineering* 45, no. 2 (2020): 331-345.

Ummenhofer, Benjamin, and Thomas Brox. “Point-based 3d reconstruction of thin objects.” In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 969-976. 2013.

Van Nguyen, Sinh, Son Thanh Le, Minh Khai Tran, and Ha Manh Tran. “Reconstruction of 3D digital heritage objects for VR and AR applications.” *Journal of Information and Telecommunication* (2021): 1-16.

Vogiatzis, George, and Carlos Hernández. “Video-based, real-time multi-view stereo.” *Image and Vision Computing* 29, no. 7 (2011): 434-441.

Whitaker, Ross T. “A level-set approach to 3D reconstruction from range data.” *International journal of computer vision* 29, no. 3 (1998): 203-231.

Wong, K-YK, and Roberto Cipolla. “Structure and motion from silhouettes.” In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV* 2001, vol. 2, pp. 217-222. IEEE, 2001.

Zhang, Zhengyou. “A flexible new technique for camera calibration.” *IEEE Transactions on pattern analysis and machine intelligence* 22, no. 11 (2000): 1330-1334.