On the Fusion of 3-D Reconstruction Techniques

Ahmed Eid and Aly Farag
Computer Vision and Image Processing Laboratory*
University of Louisville, Louisville, KY 40292
{eid,farag}@cvip.uofl.edu

Abstract – *The commercial 3-D laser scanners provide high quality reconstructions of different surfaces. However if the reconstructed surfaces are made of diffusing materials and/or exhibit self occluding properties then errors are guaranteed in the output reconstruction. In this paper, we try to rectify this situation by fusing the 3-D reconstructions generated by a 3-D laser scanner and other 3-D reconstruction techniques that may not exhibit similar reconstruction errors. We propose a fusion technique that employees a 3-D alignment criterion and a local quality assessment technique as well. The fusion decision in the proposed technique is guided by the extracted contours from the input images of the object under-reconstruction. The approach can be applied to different 3-D reconstructions from different sensors. The approach exhibits promising results on the fusion of the space carving technique and the 3-D reconstruction by a 3-D laser scanner.*

Keywords: 3-D scanner, laser, space carving, 3-D registration, evaluation, fusion.

1 Introduction

The 3-D reconstruction from sequence of images finds many applications in modern computer vision systems such as virtual reality, vision-guided surgeries, autonomous navigation, medical studies and simulations, reverse engineering, and architectural design. The very basic requirement of these applications is to find accurate and realistic reconstructions.

In fact, the 3-D scene reconstruction from multiple images is a challenging and interesting problem to tackle. It is interesting because humans naturally solve this problem in an easy and efficient way. However, it is a challenge because there is no single solution of many different solutions proposed to solve the problem has the completeness of the human's solution. Of course, there are good solutions and there may be others in the future. The current 3-D laser scanners can provide such good reconstructions. However, the laser projection/reflection is not guarantee on all surfaces especially that exhibit occlusion problems. In addition, the standard methods for extracting range data from optical triangulation scanners are accurate only for planar objects of uniform reflectance illuminated by an incoherent source. Using these methods, curved surfaces, discontinuous surfaces, and surfaces of varying reflectance cause systematic distortions of the range data. Coherent light sources such as lasers introduce speckle artifacts that further degrade the data [1].

Recently, a family of 3-D reconstruction techniques has been developed to rectify the shortcomings of existing 3-D reconstruction techniques. Examples of these techniques are: voxel coloring and space carving [2, 3, 4]. These techniques have the following advantages over the commonly used stereo techniques: (i) they account for the occlusion problem, (ii) unlike stereo they do not put any constraints on the base distances of cameras, (iii) they provide dense reconstructions and finally (iv) they provide synthetic views of photo-realistic quality for many applications of virtual reality [2].

Space carving technique [3] assumes that there is a known bounded volume in which object of interest lies. The most common approach to represent this volume is as a regular tessellation of cubes, called voxels, in the Euclidean 3-D space [5, 6]. The algorithm detains or carves a voxel depending on whether the voxel is photo-consistent or photo-inconsistent, respectively. The final reconstruction is determined when the algorithm finds all voxels photo-consistent.

In this paper, we integrate the output reconstructions by the 3-D laser scanners and the space carving technique such that the output reconstruction contains best features of both reconstructions. To achieve this goal, we have to have an efficient 3-D registration criterion and a local quality assessment technique as well [7]. The 3-D registration criterion is required to bring the 3-D reconstructions under-fusion into full alignment. Since the fusion is only required for a finite number of subsets of the given data sets, a local-error finder and a measuring criterion are required to help the fusion technique to make the fusion decision. Brief descriptions of the alignment criterion and the assessment technique used in this paper are given in section 2.

Here we propose a novel 3-D fusion technique that uses the contours of a given 3-D object under-reconstruction to guide the fusion decision task. The fusion decision in the proposed technique is a challenging problem since there is no reference 3-D reconstruction that may guide the fusion process. The object contours extracted from the given images of that object are used to take the fusion decision, we

*This work is supported by US Army under grant DABT60-02-P-0063 and Air Force Office of Scientific Research (AFOSR) grant F49620-01-1-0367.

call them the Ground Truth Contours (GTC). Similar contours of the given object are also extracted from the 3-D reconstructions under-fusion, we call them the Measured Contours (MC). The 3-D surface patches that have the closest MC to the corresponding GTC are selected in the final 3-D reconstruction.

The proposed methodology is applied to the reconstructions generated by a 3-D laser scanner and the space carving technique. The integration of such reconstruction techniques is motivated by our concern of building high quality dense 3-D reconstruction database to be used as ground truth for further studies in the performance evaluation of 3-D reconstruction techniques [7]-[14].

This paper is organized as follows. Section 2 describes the 3-D registration criterion and the quality assessment as pre-requests for the 3-D fusion process. The proposed fusion methodology is presented in section 3. Section 4 provides the results of applying the proposed methodology to the reconstructions by the space carving technique and the 3-D laser scanner. Section 5 provides the discussion and the conclusions.

2 3-D Fusion Pre-requists

A 3-D laser scanner and a CCD camera mounted on a metal arm of multiple joints attached to the scanner head are used to capture both range and intensity data [15, 16]. A mono-color, usually black, screen is attached to the scanner head facing the CCD camera such that the screen appears as a fixed background to the object under reconstruction. The structure of the mono-color screen as well as the motion mechanism of the scanner facilitate the object segmentation task. The shaft over which the scanner head is mounted, is controlled in terms of speed and angle of rotation to capture images $I_1, I_2, ...I_N$ at specific locations on a circular path.

The calibration process [17]-[20] attempts to estimate the camera parameters at each point in the circular path where the images are acquired. The camera is calibrated at the initial position to determine the projection matrix $\mathbf{P_1}$. To get full alignment of the 3-D scanner data set $\mathcal{G}$ and 3-D data set $\mathcal{M}$ generated by a given 3-D reconstruction technique, an efficient and simple 3-D registration technique is used.

2.1 3-D Data Registration

In most cases the measured data could be corrupted or of unknown quality, therefore the conventional 3-D registration techniques might not succeed to solve the problem unless human intervention is assumed. As a result another technique is needed to solve the problem. We proposed a 3-D registration technique that uses silhouettes, we called it Registration Through Silhouettes (RTS) [7]. A brief description of the RTS technique is presented in the following two paragraphs.

Since the 3-D reconstruction $\mathcal{M}$ is obtained by applying a 3-D reconstruction technique to calibrated sequence of images, a set S_{in} of silhouettes can be generated. In addition, we use the 3-D reconstruction $\mathcal{G}$ generated by the 3-D laser scanner to generate another set $\mathcal{S}_\mathcal{G}$ of silhouettes at same views as set S_{in}. In the ideal case when $\mathcal{M}$ and $\mathcal{G}$ are initially registered, S_{in} and $S_\mathcal{G}$ are aligned. However, in most cases a certain transformation $\mathcal{T}$ is needed to align $\mathcal{G}$ with $\mathcal{M}$. Applying $\mathcal{T}$ iteratively to $\mathcal{G}$ to get $\mathcal{S}_\mathcal{G}$ such that the error between S_{in} and $\mathcal{S}_\mathcal{G}$ is minimal will lead to getting the best $\mathcal{T}$ that brings $\mathcal{G}$ and $\mathcal{M}$ into match.

This registration step is mandatory if local quality assessment is required. The RTS technique has a unique feature of not using the actual 3-D reconstruction $\mathcal{M}$, hence the registration results are independent of the errors in the actual reconstruction and consequently it permits clean fusion process.

2.2 Local Quality Assessment of 3-D Reconstructions

Since $\mathcal{M}$ and $\mathcal{G}^{'} = \mathcal{T}(R,t)\mathcal{G}$ has been aligned to each other, a local quality assessment methodology is used to measure the similarity between the corresponding surface patches of both reconstructions [7]. We assign a quality index Q for each patch. The values of Q range from 0-1 with values close to 1 indicate high similarities and near-zero values indicate low similarities. A fusion decision should be taken when the values of Q are low. The lower values of Q indicate that the quality of the surface patches of one of the reconstructions should be better than the other. Which is better? and which of these patches should be retained in or eliminated from the final reconstruction? are among the questions that should be answered by the proposed fusion criterion. A brief description of the local quality assessment methodology is given below.

The reconstruction space of the given data is uniformly divided into sets of patches or voxels. Each voxel or patch j is assigned a quality index Q_j as:

$$Q^j = \frac{2R_D^j}{(R_D^j)^2 + 1}[1 - \frac{C_d^j}{C_{max}}] \tag{1}$$

where:

- R_D^j is the ratio of the deviations of the 3-D points from the centroids of patches j in the corresponding reconstructions,
- C_d^j is the Euclidean distance between the centroids of patches j in the corresponding reconstruction
- C_{max} is a constant represents the maximum possible distance between 3-D points inside patches.

If we have a reference 3-D reconstruction and other competitive 3-D reconstructions, then the final 3-D reconstruction is the composition of the patches of highest similarity index relative to the reference reconstruction. However, if we do not have a reference reconstruction, then the dissimilar patches, that have lower values of Q, should be examined to elect best patches. The proposed fusion criterion employees a decision criterion based on a closest contour method to select the closest patches to the desired reconstruction.

3 Closest Contour (CC) 3-D Fusion Methodology

Assume that there are two 3-D reconstructions Ω_1 and Ω_2 that are fully aligned and a quality index per patch/voxel is assigned to each pair of patches of the two reconstructions. Assume that one of the given reconstructions is derived from a set I of calibrated images of cardinality N.

A sequence of pre-processing techniques such as image segmentation, filtration and edge detection are applied to the set I to generate a set $\mathcal{C}$ of contour images defined as:

$$\mathcal{C} = \{c^l : c^l \subset \mathcal{C}, l = 1, ..., N\},\ \omega^l \subset c^l \tag{2}$$

where ω^l is the contour at view l. Using the projection matrix at each view as sets I and $\mathcal{C}$, the two reconstructions Ω_1 and Ω_2 are projected to the same view and the generated silhouette images are then processed to generate the contour images $\mathcal{C}_1$ and $\mathcal{C}_2$ where

$$\mathcal{C}_1 = \{c_1^l : c_1^l \subset \mathcal{C}_1, l = 1, ..., N\},\ \omega_1^l \subset c_1^l \tag{3}$$

and

$$\mathcal{C}_2 = \{c_2^l : c_2^l \subset \mathcal{C}_2, l = 1, ..., N\},\ \omega_2^l \subset c_2^l \tag{4}$$

respectively.

A synthetic set of images $\mathcal{K}$ is generated such that

$$\mathcal{K} = \{k^l : k^l \subset \mathcal{K}, (\omega^l \cup \omega_1^l \cup \omega_2^l) \subset k^l, l = 1, ..., N\} \tag{5}$$

and

$$k^l(x,y) = \begin{cases} \mathcal{L}_1, & \text{if } c^l(x,y) \in \omega^l; \\ \mathcal{L}_2, & \text{if } c_1^l(x,y) \in \omega_1^l; \\ \mathcal{L}_3, & \text{if } c_2^l(x,y) \in \omega_2^l; \\ \mathcal{L}_4, & \text{otherwise.} \end{cases} \tag{6}$$

where $\mathcal{L}_1$, $\mathcal{L}_2$, $\mathcal{L}_3$, and $\mathcal{L}_4$ are different gray levels, $1 \leq x \leq N_h$, $1 \leq y \leq N_w$, and $N_h \times N_w$ is the cardinality of k^l.

Each image k^l is uniformly divided into image windows W_j^l, where $j = 0, ..., N_m - 1$ and N_m is the number of windows W_j^l in image k^l. A number $N_c < N_m$ of windows W_j^l that contain the contour subsets ξ_ω^j, $\xi_{\omega_1}^j$, and $\xi_{\omega_2}^j$, as shown in Fig. 1a, are elected for the closest point and the closest contour tests as follows.

3.1 The Closest Point Test

For each point $p_w^i \in \xi_\omega^j$, $i = 1, ..., card(\xi_\omega^j)$ where $card(\xi_\omega^j)$ is the cardinality of ξ_ω^j, the closest points, as in [21, 22], $pc_{w_1}^i$ and $pc_{w_2}^i$ are calculated as:

$$pc_{w_1}^i = pc_{w_1}^{r_1} \tag{7}$$

such that

$$d(p_w^i, p_{w_1}^{r_1}) = \min_{h \in \{1, ...card(\xi_{\omega_1}^j)\}} d(p_w^i, p_{w_1}^h) \tag{8}$$

and

$$pc_{w_2}^i = pc_{w_1}^{r_2} \tag{9}$$

such that

$$d(p_w^i, p_{w_2}^{r_2}) = \min_{h \in \{1, ...card(\xi_{\omega_2}^j)\}} d(p_w^i, p_{w_2}^h) \tag{10}$$

where d denotes the Euclidean distance.

3.2 The Closest Contour Test

To determine which of $\xi_{\omega_1}^j$ and $\xi_{\omega_2}^j$ is closer to ξ_ω^j, the average distances $d_{av}(\xi_\omega^j, \xi_{\omega_1}^j)$ and $d_{av}(\xi_\omega^j, \xi_{\omega_2}^j)$ are calculated as:

$$d_{av}(\xi_\omega^j, \xi_{\omega_1}^j) = \frac{1}{card(\xi_\omega^j)} \sum_{i=1}^{card(\xi_\omega^j)} d_i(p_w^i, pc_{w_1}^i) \tag{11}$$

and

$$d_{av}(\xi_\omega^j, \xi_{\omega_2}^j) = \frac{1}{card(\xi_\omega^j)} \sum_{i=1}^{card(\xi_\omega^j)} d_i(p_w^i, pc_{w_2}^i) \tag{12}$$

then the closest contour segment is the one that has the minimum of $d_{av}(\xi_\omega^j, \xi_{\omega_1}^j)$ and $d_{av}(\xi_\omega^j, \xi_{\omega_2}^j)$.

It is important to note that other methods that can extract information about the shape of the contours, e.g. gradient methods, can be used to determine the closest contour in addition to the closest Euclidean distance as described above.

3.3 The Fusion Decision

In 3-D space, the surface segments accompanying to the the 2-D contours are determined during the object projection phase. Therefore the surface segment corresponding to the closest contour is already known in the 3-D space. As shown in Fig. 1b, the 3-D segment $\Xi_{\Omega_1}^0$ is accompanying to the closest contour segment $\xi_{\omega_1}^0$ in Fig. 1a. A cubic voxel V_0 that has the the same centroid as $\Xi_{\Omega_1}^0$ is constructed to include $\Xi_{\Omega_1}^0$. The surface patches/voxels inside V_0 are elected from the 3-D reconstruction Ω_1 to be in the final reconstruction. The process is repeated for each contour and surface segment to reconstruct the final output.

4 Experimental Results

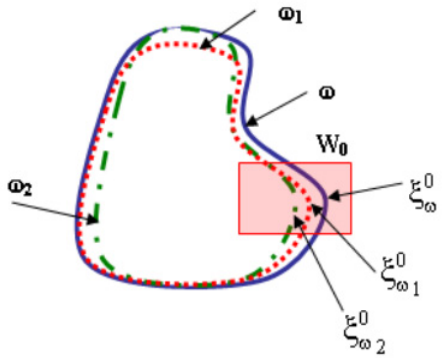
An experiment is performed on the fusion of the 3-D reconstructions by the 3-D scanner and the space carving technique. Fig. 2a shows screen captures of the 3-D reconstruction by the 3-D scanner for a bear object. Projecting sharp and thin laser lines on the surface of this bear is not guaranteed, hence the errors in the output reconstruction. Filling gaps in the surface can fix some errors in smoothed parts, the back of the bear, while for complex parts that have discontinuities, the front of the bear, the filling does not provide significant enhancements as shown in Fig. 2b.

The space carving technique is applied to a number of 12 images of the bear. A sample of the input images is shown in Fig. 3a. A ground truth silhouette image extracted from Fig. 3a is shown in Fig. 3b. The measured silhouette images at the same view as in Fig. 3b extracted by projecting the 3-D reconstructions by space carving and the 3-D laser scanner are shown in Fig. 3c and Fig. 3d, respectively. The measured silhouettes indicate differences between the reconstructions by space carving and the 3-D laser scanner.

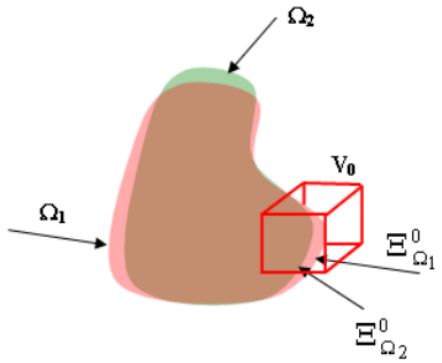
Fig. 4a shows a contour image for the ground truth contour (GTC), white, extracted from the silhouette image in Fig. 3b and a measured contour (MC), black, extracted from the silhouette in Fig. 3c. A similar image is shown

in Fig. 4b, however the MC is extracted from the silhouette in Fig. 3d. These contour images give a clue about which reconstruction at this view and at specified surface segment is closest to the desired reconstruction. Some contour segments at the back of the bear in Fig. 4a and b show that the reconstruction by the 3-D scanner is closer to the desired reconstruction, however at the top of the bear's head the reconstruction by the space carving is the closer.

The proposed 3-D fusion technique is applied to the given reconstructions of the bear by the 3-D scanner, shown in Fig. 2a, and by the space carving shown in, Fig. 5a. The fusion results are shown in Fig. 5b. As shown in the Figure the fusion process can enhance the 3-D reconstruction of a given object by selecting well-reconstructed surface segments from each reconstruction and integrate them into one reconstruction. It is important to note that for objects that have concavities, the fusion decision can not be taken based on the closest contour method. This represents a limitation of our proposed technique.



(a)

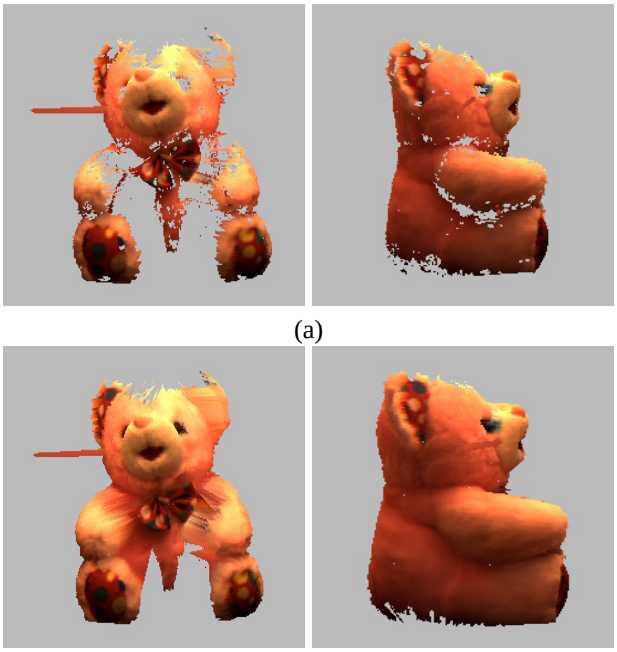


(b)

Fig. 1: (a) the ground truth contour (GTC) and the measured contours (MC) form two different reconstructions, and (b) the 3-D reconstructions at certain view from which the above contours in Fig. 1a are extracted.

5 Conclusions

Many techniques have been proposed in the literature to achieve better 3-D reconstructions for 3-D scenes. However, there is no single approach that is general for all types of surfaces and materials. Even the commercial 3-D laser scanners fail to reconstruct surfaces of self-occluding and/or diffusing properties. In this paper, we use recent volumetric approaches for 3-D reconstruction to rectify some errors introduced by the commercial 3-D laser scanners. We propose a 3-D fusion methodology that takes the fusion decision based on the presented novel concept of using the object contour to guide the fusion process. The approach is efficient and serves our goal of building reference 3-D models of challenging objects and materials that are difficult to be reconstructed using single 3-D reconstruction technique. The presented results show that the approach is promising and can be applied to different reconstructions using different 3-D modeling techniques.



(a)

(b)

Fig. 2: Screen captures of a 3-D reconstruction by 3-D laser scanner: (a) without filling the gaps (b) after filling the gaps.

References

[1] B. Curless and M. Levoy,"Better optical triangulation through spacetime analysis," *Proceedings of 5th International Conference on Computer Vision*, Boston, MA, 20-23 June 1995.

[2] S. Seitz and C. Dyer, "Photorealistic Scene Reconstruction by Voxel Coloring," *Proceedings of Computer Vision and Pattern Recognition Conference*, Puerto Rico, pp. 1067-1073, June 1997.

[3] K. Kutulakos and S. Seitz. "Theory of Shape by Space Carving," *Proceedings of IEEE International Conference On Computer Vision*, Corfu, Greece, pp. 307-314, Sept. 1999.

[4] W. Culbertson, T. Malzbender, and G. Slabaugh, "Generalized Voxel Coloring,"*International Workshop on Vision Algorithms*, pp.100-115, Greece, 1999.

[5] C.R. Dyer, "Volumetric Scene Reconstruction from Multiple Views", In L.S. Davis, editor, Foundations of Image Understanding, pp. 469-489. Kluwer, Boston, 2001.

[6] G. Slabaugh, B. Culbertson, T. Malzbender, and R. Schafer, "A survey of methods for volumetric scene reconstruction from photographs", In K. Mueller and A. Kaufmann, editors, Proceedings of the Joint IEEE TCVG and Eurographics Workshop (VolumeGraphics-01), pp. 81-100, Wien, June 2001.

[7] A. Eid and A. Farag, "A Unified Framework for Performance Evaluation of 3-D Reconstruction Techniques from Sequence of Images," *Proceedings of IEEE Workshop on Real Time 3-D Sensors and their Use, in conjunction with CVPR conference*, Washington DC., June 2004.

[8] A. Eid, A. Farag, "A Design of an Experimental Setup for Performance Evaluation of 3-D Reconstruction Techniques from Sequence of Images," *Proceedings of Applications of Computer Vision Workshop in conjunction with ECCV04* Prague, May 2004.

[9] R. Szeliski, "Prediction Error as a Quality Metric for Motion and Stereo," *Proceedings of IEEE International Conference On Computer Vision*, Corfu, Greece, pp. 781-788, 1999.

[10] R. Szeliski and R. Zabih, "An Experimental Copmarison of Stereo Algorithms," *International Workshop on Vision Algorithms*, pp. 1-19, Greece, 1999.

[11] A. Eid and A. Farag, "On the Performance Characterization of Stereo and Space Carving," *Proceedings of Advanced Concepts for Intelligent Vision Systems*, Ghent, Belgium, pp. 291-296, 2003.

[12] D. Scharstein and R. Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms," *International Journal for Computer Vision*, 47(1):7-42, 2002.

[13] http://cat.middlebury.edu/stereo/data.html

[14] J. Mulligan, V. Isler, and K. Daniilidis, "Performance Evaluation of Stereo for Tele-presence," *Proceedings of IEEE International Conference On Computer Vision*, Vancouver, Canada, Vol. II pp. 558-565, 2001.

[15] A. Eid, S. Rashad and A. Farag, "A General Purpose Platform for 3D Reconstruction from Sequence of Images," *Proceedings of 5th International Conference on Information Fusion*, Annapolis, MD, Vol. 1, pp. 425-413, July 2002.

[16] A. Eid, S. Rashad and A. Farag, "Validation of 3D Reconstruction from Sequence of Images," *Proceedings of the International Conference on Signal Processing, Pattern Recognition, and Applications SSPRA* Greece, June 2002.

[17] L. Robert. "Camera Calibration without feature extraction," *Computer Vision and Image Understanding*, 63(2), March 1996.

[18] R. Tsai, "An Efficient and Accurate Camera Calibration Technique for 3d Machine Vision," *Proceedings of IEEE Coonference on Computer Vision and Pattern Recognition*, pp. 364-374, Miami Beach, FL,1986,

[19] R. Hartley and A. Zisserman, *Multiple View Geometry in computer vision*, Cambridge University Press., 2000.

[20] O. Faugeras and Q. Luong, *The Geometry of Multiple Images*, The MIT Press., 2001.

[21] M. Ahmed, S. Yamany, and A. Farag, "Fast Algorithm for Registration of FreeForm Curves and Surfaces," IEEE International Conference on Image Processing (ICIP97), Santa Barbara, CA, Oct. 1997.

[22] P. Besl and N. McKay, " A method for registration of 3-D shapes," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 14, No 2, pp. 239-256, 1992.

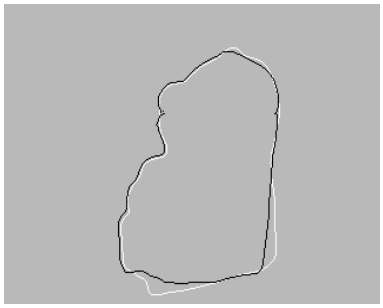
(a)

(b)

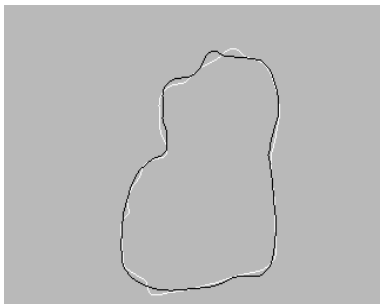
(c)

(d)

Fig. 3: (a) an example of input image to the space carving technique, (b) a silhouette image extracted from the image in Fig. 3a, (c) a silhouette image extracted from the reconstruction by space carving at the same view as in Fig. 3b, and (d) a silhouette image extracted from the reconstruction by a 3-D laser scanner at the same view as in Fig. 3b.

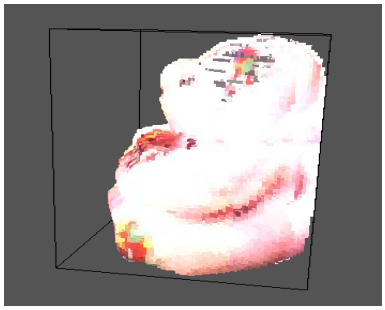


(a)

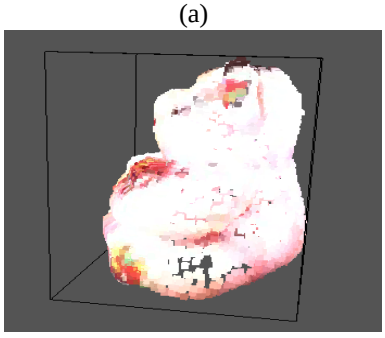


(b)

Fig. 4: (a) a contour image shows the ground truth contour (GTC), white, extracted from the silhouette image in Fig. 3b and a measured contour (MC), black, extracted from the silhouette in Fig. 3c and (b) same image as in Fig. 4a however the MC is extracted from the silhouette in Fig. 3d.



(a)



(b)

Fig. 5: (a) the reconstruction by space carving, and (b) the fusion of the scanner and space carving reconstructions based on the closest contour method.