iModel: Interactive Co-segmentation for Object of Interest 3D Modeling

Adarsh Kowdle[1], Dhruv Batra[2], Wen-Chao Chen[3], and Tsuhan Chen[1]

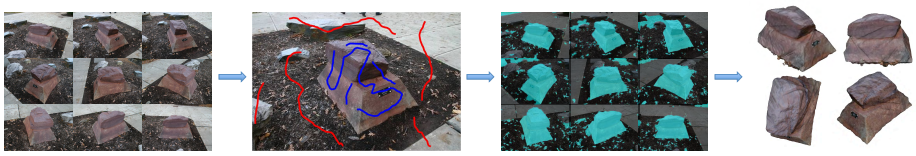
[1] Cornell University, Ithaca, NY, USA
[2] Carnegie Mellon University, Pittsburgh, PA, USA
[3] Industrial Technology Research Institute, Taiwan
apk64@cornell.edu, batradhruv@cmu.edu,
chaody@itri.org.tw, tsuhan@ece.cornell.edu

Abstract. We present an interactive system to create 3D models of objects of interest in their natural cluttered environments.
A typical setting for 3D modeling of an object of interest involves capturing images from multiple views in a multi-camera studio with a mono-color screen or structured lighting. This is a tedious process and cannot be applied to a variety of objects. Moreover, general scene reconstruction algorithms fail to focus on the object of interest to the user. In this paper, we use successful ideas from the object cut-out literature, and develop an interactive-cosegmentation-based algorithm that uses scribbles from the user indicating foreground (object to be modeled) and background (clutter) to extract silhouettes of the object of interest from multiple views. Using these silhouettes, and the camera parameters obtained from structure-from-motion, in conjunction with a shape-from-silhouette algorithm we generate a texture-mapped 3D model of the object of interest.

1 Introduction

If there is one thing the growing popularity of immersive virtual environments (like Second-Life® with 6.1 Million members) and gaming environments (like Project Natal®) has taught us – it is that people crave personalization. For example, gamers want to be able to "scan" and use their own gear (like skateboards) in a skateboarding game. While there exist some tools to enable this implanting of real-world objects in virtual environments, we believe this is an important problem, worth studying formally by computer vision researchers. This paper takes a first step towards enabling users to create 3D models of an object of interest, which may then be easily implanted in a virtual environment.

One approach to achieve this, would be to haul an expensive laser scanner to get precise depth estimates in a controlled setup, and reconstruct the object [1]. However, this might be not be a feasible solution for average users. Another typical approach for this problem is to capture images of the object in a controlled environment like a multi-camera studio with mono-color screen and structured lighting, and use a shape-from-silhouette algorithm [2–5] to render the 3D model. Although these techniques have produced promising results in these constrained settings, this is a tedious process, and in some cases not an option (for example, immovable objects like a statue, historically



(a) (b) (c) (d)

Fig. 1: Overview of system: (a) Stone dataset (24 images) - subset of images given to the system shown; (b) User interactions to indicate the object of interest (blue scribbles = object of interest, red scribbles = background); (c) Resulting silhouettes after co-segmentation (object of interest in cyan color) (d) Some sample novel views of the rendered 3D model (Best viewed in color).

or culturally-significant artifacts). However, in a world where we are surrounded by cellphone-cameras, a more accessible approach is to capture images of the object in its natural environment and directly estimate the 3D structure from these natural images. The images captured in this case would typically have cluttered backgrounds, which is known to be problematic for background subtraction algorithms.

Overview. In this work, we present an interactive system to create 3D models of objects of interest in their natural cluttered environments. Our approach builds on the success of recent works in interactive co-segmentation [6–9] that have shown that foreground and background statistics can be shared across images to jointly segment or *co*-segment the common foreground from multiple images. Our interactive algorithm, uses scribbles from the user indicating foreground and background to extract silhouettes of the object of interest from multiple views. Using these silhouettes, and camera parameters obtained from structure-from-motion [10], in conjunction with a octree-reconstruction-based shape-from-silhouette algorithm [2, 4] we generate a texture mapped 3D model of the object of interest. We demonstrate the effectiveness of our algorithm on a wide range of objects and show that this same approach extends to obtain 3D models of even monuments.

Contributions. The main contribution of this paper is a simple interactive system for 3D modeling of an *object of interest*. The approach obviates the need for a complicated, controlled studio environment and works for objects in their natural cluttered environments. To the best of our knowledge this is the first approach to use ideas from the interactive co-segmentation literature for object of interest 3D modeling.

Organization. The rest of this paper is organized as follows. Section 2 discusses related work. Section 3 describes our approach to co-segment the object of interest in all images using our interactive algorithm and then extract a 3D model of this object of interest using these silhouettes. Section 4 presents our modeling results on a wide range of objects. Finally, we conclude in Section 5 with discussions.

2 Related Work

Controlled Setups. Several works [11–16] use a multi-camera studio setup, with controlled lighting and a mono-color screen to capture images. This allows for easy back-

ground subtraction with chroma keying. A shape-from-silhouette algorithm can be applied to the silhouettes obtained after background subtraction [2–5]. Levoy et al. [1] construct Cyberware gantry, a laser scanner setup that is able to obtain extremely precise depth map of the object. Zhang et al. [17] perform spacetime analysis by sweeping multiple color stripes across the object to obtain the shape of the object. They also propose an approach of using a structured lighting to perform spacetime stereo by matching a pair of video streams which can help recover the shape of even dynamic objects [18]. Yezzi et al. [19] propose 'stereoscopic segmentation' to obtain the silhouettes of the object, which works well in a controlled setting of an object with a lambertian surface and constant albedo. Lee et al. [20] also propose a method to obtain silhouettes of the foreground object with the assumption that the background is homogeneous to some extent, and differs from foreground. We note that all these methods require a controlled setup and are not (directly) applicable to a wide range of objects that cannot be captured in such a controlled enviroment.

Multiview Stereo. When we move from images taken in controlled environments to images taken outdoors in cluttered scenes, background subtraction becomes a significantly harder task. In these cases, algorithms typically try to reconstruct the entire scene as a whole and do not focus on any object of interest. A popular work by Snavely et al. [21] relies on structure-from-motion to render sparse point clouds. However, this results in only a sparse reconstruction and not a complete 3D model of the object, which is the goal of this paper. Multiview stereo algorithms [22] like patch-based multi-view stereo introduced by Furukawa et al. [23] can generate reasonably dense models. Other notable dense-reconstruction algorithms include Van Gool et al. [24] which offers dense 3D reconstruction from user-supplied images via a publicly available web service. Goesele et al. [25] and Furukawa et al. [26] worked with internet-scale community photo collections. They used a multi-view stereo approach to get dense reconstructions of the geometry of objects like monuments and statues using many images. We note that our scenario is slightly different from these works – we focus on the 3D reconstruction of the object *of interest* alone and typically have access to a few consumer images (significantly fewer than community photo collections).

Interactive Algorithms. As discussed before, there have been a number of automatic algorithms to obtain the silhouettes of the object. However, the notion of an 'object of interest' clearly requires some form of user input. This is especially true when the natural surroundings of the object being modeled make it difficult to automatically extract the object from the background (see for example, the stone dataset in Figure 2a). Campbell et al. [27] tried to incorporate user interaction by assuming that the object of interest is at the center of attention. Thus, they used a seed at the center of the image to perform region growing to extract the foreground across images. This would work when the object has a fairly uniform color but, would fail if the object had multiple colors. Also, this approach does not allow the algorithm to recover from an incorrect labeling. We believe that our application requires a fully interactive algorithm. This was first explored in image segmentation by Boykov and Jolly [28] who posed interactive segmentation as a discrete optimization problem. Li et al. [29] and Rother et al. [6] presented simplified user interactions and other improvements to the basic framework. This idea has been extended to object modeling by Sormann et al. [30]. They start with

(a) (b) (c)

Fig. 2: Interactive Co-segmentation: (a) Group of images; (b) User interactions to indicate the object of interest (blue scribbles = object of interest, red scribbles = background); (c) Resulting silhouettes after co-segmentation (in cyan color). For this example, we only needed to scribble on a single image, but our system allows for scribbles on multiple images. Other groups in our dataset with more diverse appearances needed scribbles on multiple images. (Best viewed in color)

a dense binary segmentation of one of the images using intelligent scrissors or grab-cut. The foreground and background color model are then learnt using this segmentation and this model is then tranfered to the other views. The problem with this approach is that the background of the image can change very rapidly and a segmentation on one image will not be representative of this. We use recent work by Batra et al. [9] which extends this interactive approach to co-segmentation of groups of images. This makes for a very simple interactive system and also allows the user to provide additional interactions if required.

Hengel et al. [31] and Sinha et al. [32] have proposed interactive 3D modeling systems that produce piecewise planar approximations. Our work differs from them, in that we make no such planar assumptions and can model even non-planar objects.

3 Approach

We now describe our approach. We are given multiple images of the object of interest taken in cluttered scenes. Our approach, involves two main parts: 1) an energy-minimization framework for interactive co-segmentation that extracts silhouettes of the object of interest and 2) a shape-from-silhouette approach to obtain the final texture-mapped 3D model. An overview of the system is shown in Fig. 1.

3.1 Interactive Co-segmentation

We create a simple interface to accept interactions from the user in the form of scribbles to indicate the object of interest (foreground) and background as shown in Fig. 2b. Our interactive co-segmentation approach is based on the work of Batra et al. [9], which we briefly describe here.

We cast our co-segmentation problem as a binary labeling energy-minimization problem solved via graph-cuts. We begin by over-segmenting each image. The task

is to label each superpixel[4] in each image as foreground (object of interest) or background. For each image, we construct a graph over superpixels, where adjacent superpixels are joined by an edge. Associated with each graph is an energy which is a weighted combination of a data-term and an edge-term. We model the data-term as the negative log-likelihood of the features extracted at superpixels given the class model. Our features are average Luv colour features extracted over superpixels, and the class model is a Gaussian Mixture Model, which is learnt from *all* labeled superpixels (in all images). We consider a superpixel labeled if any pixel within it has been scribbled on. Our approach allows for scribbles on single or multiple images. The edge-term is a contrast sensitive Potts model. Finally, we use one graph-cut per image to compute the MAP labels for all superpixels, using the implementation provided by Bagon [34] and Boykov et al. [35, 36] and Kolmogorov [37]. This results in a co-segmentation of the object of interest across multiple images. More details about the performance of this co-segmentation algorithm may be found in Batra et al. [9].

3.2 Shape from Silhouette

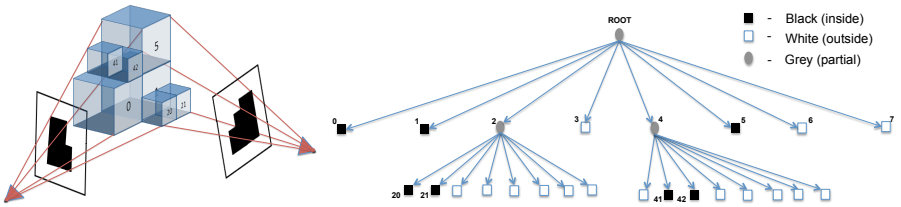
We have now extracted the silhouettes of the object of interest from multiple views. We use the structure-from-motion implementation by Snavely et al. [10] called 'Bundler' to recover camera parameters for each image. We then use a shape-from-silhouette approach to extract a volumetric 3D reconstruction of this object. Specifically, we use an octree-reconstruction method [2, 4].

An octree is a tree-structured representation which is used to describe a set of binary-valued data (in this case indicating the presence or absence of cubes of voxels in the 3D model). The octree is constructed by recursively subdividing a cube to eight sub-cubes, starting with the root which represents the bounding volume in which the object of interest lies as shown in Fig. 3. This cube representation captures the high degree of coherence between adjacent voxels. Each sub-cube in an octree is projected onto the silhouette images and can be one of three colors. A *black* node indicates that the cube is totally occupied (i.e. it projects completely inside the silhouettes), and a *white* node indicates that it is totally empty (i.e. it projects completely outside the silhouettes). Both black cubes and white cubes are leaf nodes in the tree. A *gray* node indicates that the cube lies on the boundary of the object and is only partially filled. The gray cube is subdivided till each of the sub-cubes can be assigned a black or white color.

We use a variant of this algorithm which is optimized for speed [2, 4]. For more details the reader is referred to Szeliski et al. [2] and Chen et al. [4]. We note here that our system allows users to visually examine the 3D reconstruction and give more scribbles to improve silhouettes that would in turn lead to better 3D reconstruction. However, we do not perform multiple iterations for our experiments.

It is worth mentioning that shape-from-silhouette algorithms have well-understood limitations. Specifically, they are unable to model certain concavities in the structure (e.g. details on the surface of the structure). However, as we show through our results

[4] We use mean-shift [33] to extract these superpixels, and typically break down 350×500 images into ~ 400 superpixels per image.



(a) (b)

Fig. 3: Simple two-level octree: (a) Synthetic visualization to show octree reconstruction (Silhouettes of the object shown in black in two views); (b) Corresponding tree representation of the two-level octree (Best viewed in color).

(Section 4), our approach might be a good first-approximation for consumer applications. Recently, Fabbri et al. [38] developed an algorithm which allows for reconstructing curves in the structure. Although not done here, this approach could provide cues to help tackle the problem of concavities in the structure.

4 Results and Discussions

We demonstrate the effectiveness of our algorithm on a number of datasets ranging from a simple collection taken in a controlled setup to a community photo collection and a video captured in cluttered scenes. In this section, we show the rendered 3D model for each dataset, captured from novel view-points. For all datasets except the dino dataset, we texture map the model by back-projecting the faces of the mesh onto a single image. We observed that in outdoor scenes texture mapping from multiple views can lead to some artifacts at the seams due to changes in illumination.

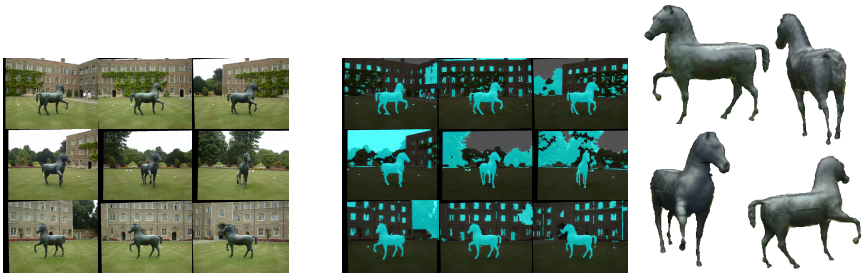
Dino Dataset. The first dataset we use is a standard dataset from the Oxford Visual Geometry Group[5], shown in Fig. 4a. One of the images in the dataset was chosen at random and the interactions were provided to indicate the dino as foreground and the blue screen as the background. The resulting silhouettes are shown in Fig. 4b. The 3D model obtained from the shape-from-silhouette algorithm. This dataset was captured in a controlled setup which allowed us to texture map the model using multiple views i.e. by projecting the faces onto the corresponding image where they are visible. Occlusion poses a significant problem for multiview texturing, we use the approach of Chen et al. [39] to overcome this by using the depth buffer (z-buffer) data from the graphics card.

This result simply serves as a proof of concept under a controlled setup, and it is encouraging to see that our approach is able to render a good reconstruction *without* any prior knowledge about this setup.

[5] Oxford Visual Geometry Group multiview dataset: http://www.robots.ox.ac.uk/~vgg/data/data-mview.html

(a) (b) (c)

Fig. 4: Dino dataset (36 images): (a) Subset of the collection of images given to the system where the dino was marked the object of interest; (b) Resulting silhouettes after co-segmentation (in cyan color); (c) Some sample novel views of the 3D model (Best viewed in color).

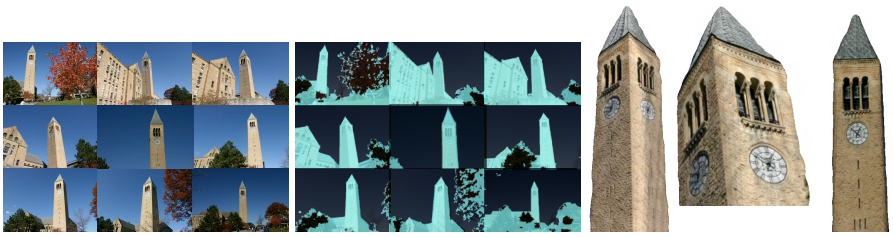


(a) (b) (c)

Fig. 5: Cambridge unicorn dataset (14 images): (a) Subset of the collection of images given to the system where the unicorn statue was marked as the object of interest; (b) Resulting silhouettes after co-segmentation (in cyan color); (c) Some sample novel views of the 3D model (Best viewed in color).

Cambridge Unicorn Dataset. We use the Cambridge unicorn dataset [40], shown in Fig. 5a. Using interactions to indicate the unicorn as the object of interest, we obtain the silhouettes as shown in Fig. 5b which results in the texture-mapped 3D model as shown in Fig. 5c.

Stone Dataset. This dataset demonstrates that our algorithm performs well even when the background becomes highly cluttered. The stone dataset is shown in Fig. 2a. Note that the stone is visually very similar to the ground surface. The silhouettes obtained after providing the interactions to indicate the stone as object of interest, are shown in Fig. 2c. It is worth mentioning that the noisy (incorrectly-labeled) superpixels in the segmentations can be removed by increasing the smoothness penalty in our energy-minimization framework. However, these sparse incorrectly-labeled superpixels do not affect the reconstruction, as they get filtered out in the shape-from-silhouette algorithm. The texture-mapped 3D model obtained from the shape-from-silhouette algorithm is



(a) (b) (c)

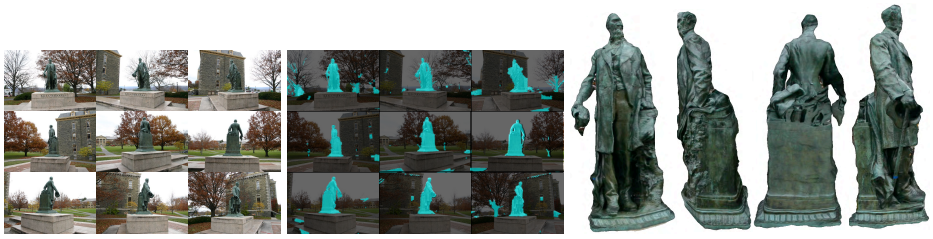
Fig. 6: Clock tower dataset (32 images): (a) Subset of the collection of images given to the system where the clock tower was marked as the object of interest; (b) Resulting silhouettes after co-segmentation (in cyan color); (c) Some sample novel views of the 3D model (Best viewed in color).

shown in Fig. 1d.

Clock Tower Dataset. We now try to reconstruct immovable objects that cannot be taken to a standard studio setup as shown in Fig. 6a. With interactions we obtain the silhouettes of our object of interest (clock tower), as shown in Fig. 6b which can be used to obtain texture mapped 3D model using the shape-from-silhouette algorithm as shown in Fig. 6c. We note that algorithms like structure-from-motion and patch based multi-view stereo would try to reconstruct the whole scene and result in an incomplete reconstruction in this case.

Statue Dataset. We now demonstrate the effectiveness of our algorithm on images where the background changes drastically as shown in Fig. 7a. The silhouettes of the object of interest (statue) obtained using our algorithm are shown in Fig. 7b. The texture mapped 3D model of the statue obtained using these silhouettes are shown in Fig. 7c. Note here that a part of the head of the statue gets clipped off in the generated model. The reason for this is a leak in the superpixels where a portion of the head became part of the sky superpixel. We can overcome this problem by working on pixels instead of superpixels i.e. set up the energy minimization over a graph of pixels instead of superpixels. This would increase the computational complexity but result in better silhouettes.

Video Dataset. Our work opens up the possibility of allowing users to render themselves as avatars in virtual worlds. We consider this scenario of reconstructing a person in 3D. We captured a video of the person to be modeled by walking around them. Selected frames this video are shown in Fig. 8a. With interactions, we obtain the silhouettes as shown in Fig. 8b which results in the texture mapped 3D model as shown in Fig. 8c. We can see that the reconstruction is fairly complete, however we observe a leak in the superpixel map here as well.



(a) (b) (c)

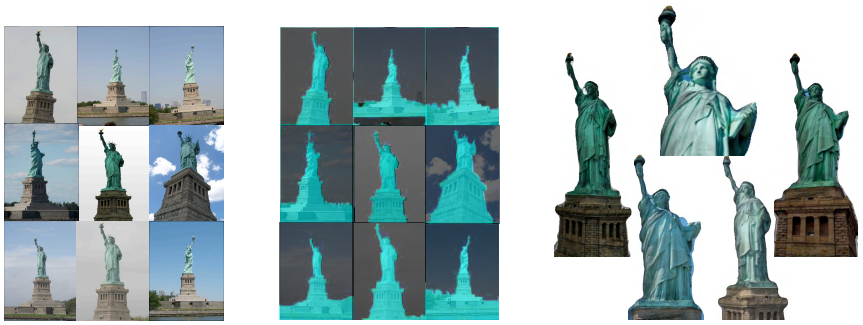
Fig. 7: Statue dataset (38 images): (a) Subset of the collection of images given to the system where the statue was marked as the object of interest; (b) Resulting silhouettes after co-segmentation (in cyan color); (c) Some sample novel views of the 3D model (Best viewed in color).

(a) (b) (c)

Fig. 8: Video dataset (17 images obtained by sampling the video): (a) Subset of the collection of images given to the system where the person was considered the object of interest; (b) Resulting silhouettes after co-segmentation; (c) Some sample novel views of the 3D model (Best viewed in color).

Community Photo collection - Statue of Liberty dataset. With millions of images available on the internet, we consider an application geared towards internet-scale reconstruction of objects where the user searches for an object of interest, in this case the *Statue of Liberty*. We start with a set of 1600 images of the Statue of Liberty collected by Snavely et al. from Flickr®. We use all the images to estimate the camera matrices using structure-from-motion [10]. For our algorithm, we sampled a subset of 15 images spanning a large field of view, as shown in Fig. 9a. The silhouettes are shown in Fig. 9b and the texture-mapped 3D model are shown in Fig. 9c. We note here that there are a few artifacts like the blue sky above the shoulder as well as the thinned arm. Some of these problems (like the superpixel leaks) may be corrected by working with pixels. However, some (like the lack of detail on the face of the Statue of Liberty) are a direct result of our reliance on a segmentation framework and may not be possible to fix. The results on this dataset have also been reported by the multi-view stereo work of Goesele et al. [25], where they obtain a dense depth model for the statue. A comparison between

the model we generate, the point cloud model from photo-tourism [10] and multi-view stereo model [25] is shown in Fig. 10.

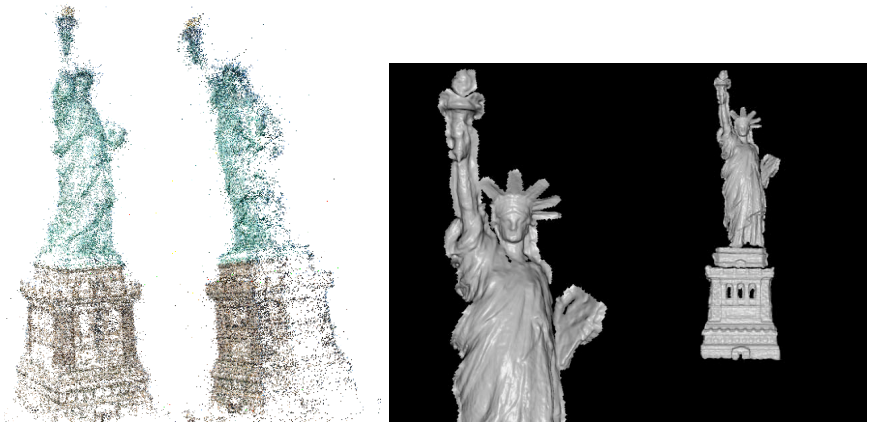


(a) (b) (c)

Fig. 9: Community photo collection - Statue of Liberty dataset: (a) Subset of the collection of images given to the system - for our co-segmentation algorithm we use a subset of 15 images spanning a large field of view from a collection of 1600 images; (b) Resulting silhouettes after co-segmentation (in cyan color); (c) Some sample novel views of the 3D model (Best viewed in color).

It is worth mentioning that once we obtain the silhouettes of the object of interest, we can use any known shape-from-silhouette algorithm at this stage to obtain the 3D model (not necessarily the octree-based reconstruction approach we used here), for example, the approach by Wong et al [41]. In addition, we can use this co-segmentation algorithm with the popular multi-view stereo reconstruction approach, to focus the output of the multi-view stereo algorithm on the object of interest. As an illustration, we show the model generated using patch-based multi-view stereo (PMVS) [6] in Fig. 11 when constrained by the silhouettes extracted using our interactive co-segmentation algorithm. We used the statue dataset in Fig. 7a for this experiment. In Fig. 11a, we show the result of PMVS without any prior knowledge of the object of interest. In Fig. 11b we show the 3D model obtained from PMVS using our silhouettes. As we explained earlier, multi-view stereo algorithms would try to reconstruct the whole scene without giving importance to the object of interest. We can see that use of silhouettes helps obtain a more accurate 3D model of the object of interest. Another crucial advantage of using the silhouettes is to speed up the multi-view stereo algorithm with geometrically consistent reconstructions. In our experiment with PMVS, it took 3 hours to obtain the model in Fig. 11a, as opposed to 8 minutes using the silhouettes to render Fig. 11b. However, faster implementations may be available for PMVS.

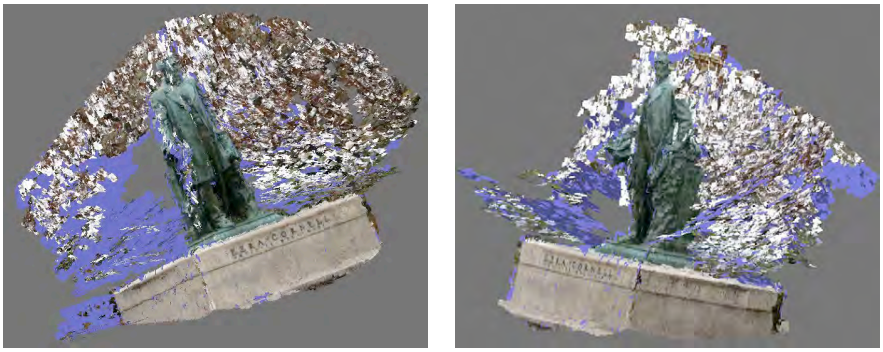
[6] We use the PMVS implementation described in [23] available at http://grail.cs.washington.edu/software/pmvs/pmvs-1/index.html



(a) (b)

(c)

Fig. 10: Statue of Liberty comparison: (a) Point cloud reconstruction by photo-tourism, using 1600 images; (b) Dense reconstruction using multi-view, using 72 images (figure from [25], used with permission). With a lot of images, multi-view stereo can give a good depth model; (c) Pleasing texture mapped reconstruction rendered using our interactive co-segmentation algorithm, using 15 images. (Best viewed in color).



(a)

(b)

Fig. 11: Patch-based multi-view stereo experiment using images in Fig. 7a where the statue is the object of interest: (a) When the silhouettes are not available PMVS tries to reconstruct the whole scene as shown; (b) Using the silhouettes produced by our co-segmentation algorithm, we can use PMVS to obtain the 3D model of the statue which was the object of interest (Best viewed in color).

5 Conclusions and Future Work

With the growing popularity of immersive virtual environments and large-scale reconstructions in mind, we present a simple interactive algorithm which enables the user to obtain a 3D model of an object of interest and render it as part of the reconstruction.

The interactive algorithm obviates the need for a complicated, controlled environment and works reasonably well in cluttered scenes. We demonstrate the effectiveness of the algorithm by modeling a wide range of objects captured in cluttered environments, *in the wild*. We also show that the same system extends well to community photo collections, thus taking a step towards building better large scale 3D environments.

We note that, in our work we only make use of camera parameters and *not* correspondences or 3D positions of feature points. As a future work, we want to incor-

porate this information which should help obtain better reconstructions. Moreover, in this work, co-segmentation and shape-from-silhouette steps were used purely in a "feedforward" manner. A possible future direction would be to place the 3D modeling and 2D image co-segmentation into an iterative loop where they aid each other. Geometric consistency constraints between different images would help achieve better co-segmentation and thereby help create better 3D models. In addition, improved techniques to use texture from multiple views while texture mapping objects in outdoor scenes would be useful and can be explored in the future.

References

1. Levoy, M., Pulli, K., Curless, B., Rusinkiewicz, S., Koller, D., Pereira, L., Ginzton, M., Anderson, S., Davis, J., Ginsberg, J., Shade, J., Fulk, D.: The digital michelangelo project: 3d scanning of large statues. In: Siggraph. (2000) 131–144
2. Szeliski, R.: Rapid octree construction from image sequences. CVGIP: Image Understanding **58** (1993) 23–32
3. Fang, Y.H., Chou, H.L., Chen, Z.: 3d shape recovery of complex objects from multiple silhouette images. Pattern Recogn. Lett. **24** (2003) 1279–1293
4. Chen, Z., Chou, H.L., Chen, W.C.: A performance controllable octree construction method. In: ICPR. (2008) 1–4
5. Forbes, K., Nicolls, F., de Jager, G., Voigt, A.: Shape-from-silhouette with two mirrors and an uncalibrated camera. In: ECCV. (2006) 165–178
6. Rother, C., Kolmogorov, V., Blake, A.: "Grabcut": interactive foreground extraction using iterated graph cuts. SIGGRAPH (2004)
7. Hochbaum, D.S., Singh, V.: An efficient algorithm for co-segmentation. In: ICCV. (2009)
8. Mukherjee, L., Singh, V., Dyer, C.R.: Half-integrality based algorithms for cosegmentation of images. In: CVPR. (2009)
9. Batra, D., Kowdle, A., Parikh, D., Luo, J., T, C.: icoseg: Interactive co-segmentation with intelligent scribble guidance. In: CVPR. (2010)
10. Snavely, N., Seitz, S., Szeliski, R.: Photo tourism: Exploring photo collections in 3d. In: SIGGRAPH. (2006) 835–846
11. Franco, J.S., Boyer, E.: Exact polyhedral visual hulls. In: BMVC. Volume 1. (2003) 329–338
12. Starck, J., Hilton, A.: Surface capture for performance-based animation. IEEE Computer Graphics and Applications **27** (2007) 21–31
13. Vlasic, D., Baran, I., Matusik, W., Popović, J.: Articulated mesh animation from multi-view silhouettes. In: SIGGRAPH, ACM (2008) 1–9
14. Curless, B., Levoy, M.: A volumetric method for building complex models from range images. In: SIGGRAPH, ACM (1996) 303–312
15. Chen, Y., Medioni, G.: Object modelling by registration of multiple range images. Image Vision Comput. **10** (1992) 145–155
16. Andrew W. Fitzgibbon, G.C., Zisserman, A.: Automatic 3d model construction for turn-table sequences. In: Proceedings of SMILE Workshop on Structure from Multiple Images in Large Scale Environments. Volume 1506. (1998) 154–170
17. Zhang, L., Curless, B., Seitz, S.M.: Rapid shape acquisition using color structured light and multi-pass dynamic programming. 3DPVT (2002) 24
18. Zhang, L., Curless, B., Seitz, S.M.: Spacetime stereo: Shape recovery for dynamic scenes. CVPR **2** (2003) 367
19. Yezzi, A., Soatto, S.: Stereoscopic segmentation. IJCV **53** (2003) 31–43

20. Lee, W., Woo, W., Boyer, E.: Identifying foreground from multiple images. In: ACCV. (2007)
21. Snavely, N., Seitz, S.M., Szeliski, R.: Modeling the world from internet photo collections. IJCV **80** (2008) 189–210
22. Seitz, S.M., Curless, B., Diebel, J., Scharstein, D., Szeliski, R.: A comparison and evaluation of multi-view stereo reconstruction algorithms. In: CVPR. Volume 1. (2006) 519–528
23. Furukawa, Y., Ponce, J.: Accurate, dense, and robust multi-view stereopsis. PAMI **32** (2010) 1362–1376
24. Vergauwen, M., Van Gool, L.: Web-based 3d reconstruction service. Mach. Vision Appl. **17** (2006) 411–426
25. Goesele, M., Snavely, N., Curless, B., Hoppe, H., Seitz, S.M.: Multi-view stereo for community photo collections. In: ICCV. (2007) 265–270
26. Furukawa, Y., Curless, B., Seitz, S.M., Szeliski, R.: Towards internet-scale multi-view stereo. In: CVPR. (2010)
27. Campbell, N., Vogiatzis, G., Hernndez, C., Cipolla, R.: Automatic 3d object segmentation in multiple views using volumetric graph-cuts. In: BMVC. (2007)
28. Boykov, Y., Jolly, M.P.: Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images. ICCV (2001)
29. Li, Y., Sun, J., Tang, C.K., Shum, H.Y.: Lazy snapping. SIGGRAPH (2004)
30. Sormann, M., Zach, C., Bauer, J., Karner, K., Bischof, H.: Automatic foreground propagation in image sequences for 3d reconstruction. In: DAGM. (2005)
31. Hengel, A., Dick, A.R., ThormŁhlen, T., Ward, B., Torr, P.H.S.: Videotrace: rapid interactive scene modelling from video. ACM Trans. Graph. **26** (2007) 86
32. Sinha, S., Steedly, D., Szeliski, R., Agrawala, M., Pollefeys, M.: Interactive 3d architectural modeling from unordered photo collections. ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia 2008) (2008)
33. Comaniciu, D., Meer, P.: Mean shift: a robust approach toward feature space analysis. PAMI **24** (2002) 603–619
34. Bagon, S.: Matlab wrapper for graph cut (2006)
35. Boykov, Y., Kolmogorov, V.: An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. PAMI **26** (2004) 1124–1137
36. Boykov, Y., Veksler, O., Zabih, R.: Efficient approximate energy minimization via graph cuts. PAMI **20** (2001) 1222–1239
37. Kolmogorov, V., Zabih, R.: What energy functions can be minimized via graph cuts? PAMI **26** (2004) 147–159
38. Fabbri, R., Kimia, B.B.: 3D curve sketch: Flexible curve-based stereo reconstruction and calibration. In: CVPR. (2010)
39. Chen, W.C., Chou, H.L., Chen, Z.: A quality controllable multi-view object reconstruction method for 3d imaging systems. JVCIR **21** (2010) 427 – 441
40. Wong, K.Y.K., Cipolla, R.: Reconstruction of sculpture from its profiles with unknown camera positions. IEEE Transactions on Image Processing **13** (2004) 381–389
41. Wong, K.Y.K., Cipolla, R.: Structure and motion from silhouettes. In: ICCV. (2001)