Computer Vision
Homework Assignment 9 : Shape from Silhouettes

Autumn 2018

Nicolas Marchal

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Computer Vision
and Geometry Lab

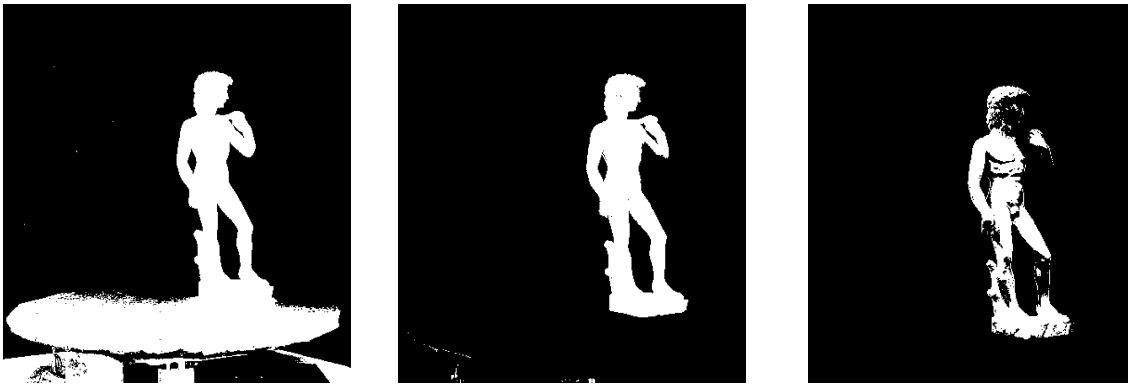
1 Visual Hull Results

For this assignment I used the following parameters

- **Silhouette Threshold:** 120
- **Bounding Box:** [minX minY minZ; maxX maxY maxZ]
  [0.3 -0.15 -1.8 ; 2.1 1.1 2.5]
- **Volume Resolution:** 128 x 128 x 256
- **volumeThreshold:** 17 (unchanged)

1.1 Silhouette Threshold

We must chose the threshold to clearly extract the silhouette from the black background. I show in figure 1 the effect of a too high or too low threshold and chose the best one : 120.



(a) Threshold = 50. Too low as we can not extract the statue from the ground

(b) Threshold = 120. Perfect as we can extract the silhouette.

(c) Threshold = 200. Too high: now the ground is completely black, but we also lose some parts of the silhouette.

Figure 1: Silhouette Threshold

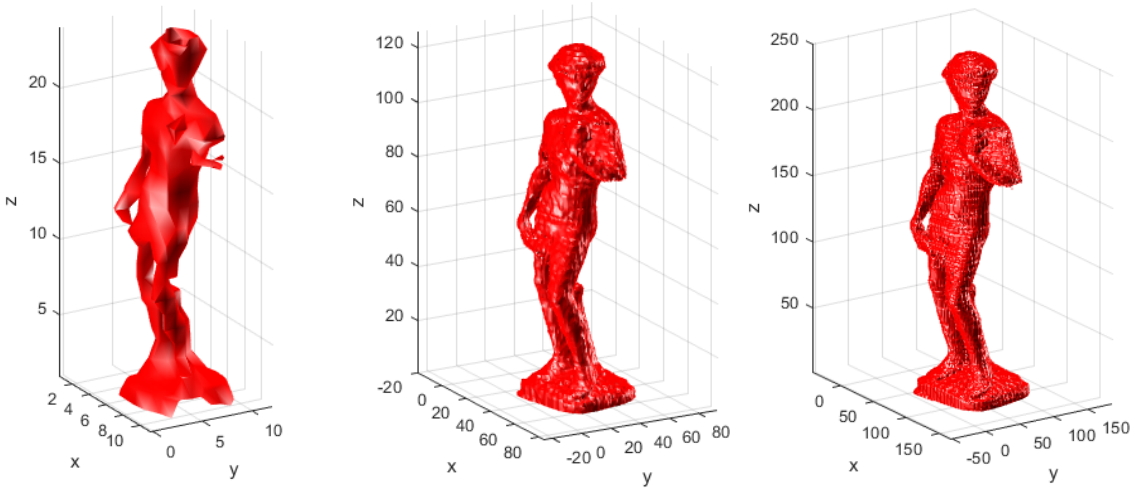
1.2 Bounding Box

The parameters influence the image like this :

- **minX:** the distance on the right of the statue
- **maxX:** the distance on the left of the statue
- **minY:** the distance in front of the statue
- **maxY:** the distance behind the statue
- **minZ:** the distance below the statue
- **maxZ:** the distance above the statue

1.3 Volume Resolution

The volume resolution influences the resolution of the reconstructed image as seen in 2. As the code is really fast to run (under a minute), we can chose the relatively high resolution of 128 x 128 x 256.

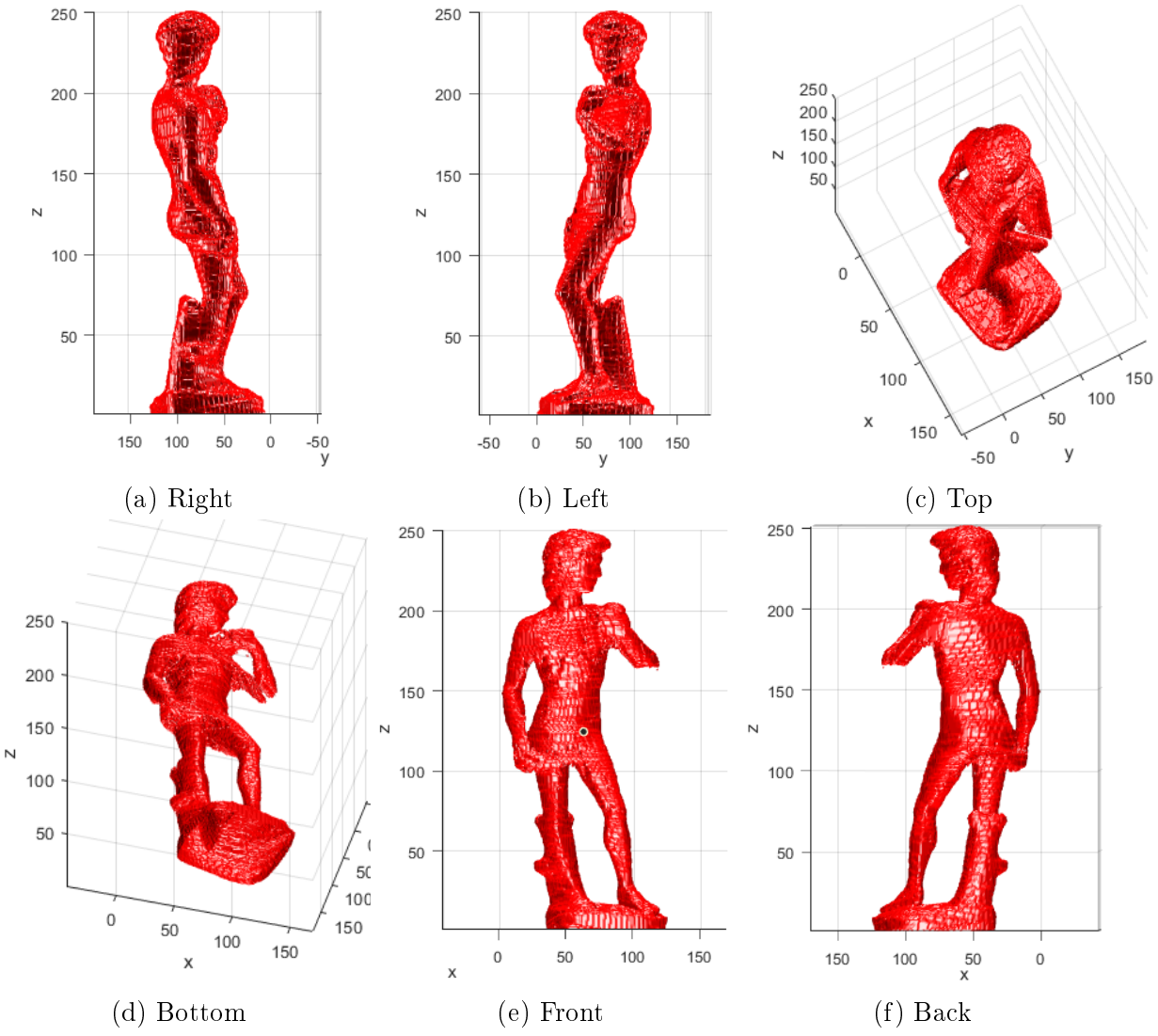


(a) 10 x 10 x 20 (b) 64 x 64 x 128 (c) 128 x 128 x 256

Figure 2: Volume resolution

2 Results

The results of my algorithm with the parameters defined earlier are shown below in figure 3



(a) Right (b) Left (c) Top

(d) Bottom (e) Front (f) Back

Figure 3: Volume resolution

3 Discussion

What is the main drawback of the shape from silhouettes approach ?

First of all, this method assumes that we can preciely extract the silhouette from the background. This assumptions is satisfied in well controlled environment like the lab where the statue was taken, but it is usually not satisfied with general images. If the silhouette can not be extracted from the background, this method will fail to reconstruct the shape. With modern approach using convolutional neural networks (CNN), the segmentation of the silhouette and the background can probably be improved, thus improving visual hull techniques.

Secondly, the intersections of all the silhouette cones tend to give unsmooth surfaces, especially if the number of cameras is limited. To smoothen the image we could interpolate the resulting 3D shape with a smoothing function.

In addition, the visual hull only provides an upper bound for the shape of an object, not the actual size. If we lack of different images of an object, this upper bound might be considerably bigger than the object and not optimal for task such as obstacle avoidance.

Finally, the visual hull does not provide good results when more than one silhouette is in the image. This can lead to ghost objects (figure 4). When silhouette overlap, it is also difficult to distinguish them. [2] proposes a solution to help with multiple silhouettes.

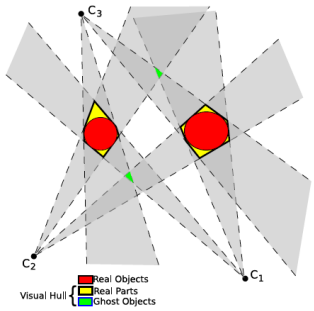


Figure 4: Ghost Objects

Is there additional information in the images that could be used?

The only was to obtain better shapes using visual hull is to increase the number of Silhouette. We can have more silhouette by increasing the number of cameras, but this is often not practical (more expensive, system setup limitation etc). Instead of increasing the number of cameras, [1] suggests to increase the number of images using multiple silhouette images across time. As it is explained in [1], there generally are an infinite number of consistent rigid motion/object shape pairs which produce the same sets of silhouette images at two different time instants. This would make the alignment between two visual hull ambiguous. However, shape matching does not make use of the **color in the images**, which is a very useful information. In [1] they actually used the color to break the alignment ambiguity.

The color intensity will be linked to the **lightening of a scene** and the **shade**. Using these can add information in our image, but it might be challenging if some pictures are taken outside under different weather conditions or different time of day.

References

[1] G. K. Cheung, S. Baker, and T. Kanade. Visual hull alignment and refinement across time: A 3d reconstruction algorithm combining shape-from-silhouette with stereo. Technical report, 2003.

[2] B. Michoud, E. Guillou, H. M. Briceno, S. Bouakaz, et al. Silhouettes fusion for 3d shapes modeling with ghost object removal.