ETH ZURICH, D-INFK

Computer Vision Exercise 5: Shape from Silhouettes

Xingchang Huang

November 23, 2017

1 SILHOUETTE EXTRACTION

First we need to extract the silhouette from given statue image by a thresholding technique so that the silhouette of the statue can be clearly extracted. Here is the problem of how to decide the threshold. In this image with clear foreground and background, we just need to easily set the value of gray-scale of pixels to be the threshold. I have tried some possible solutions of threshold including 95, 105, 120 and found that lower threshold here can extract more information of this statue while higher threshold can better filter the background information. The pictures compare the result with threshold 95 and 120 in extracted silhouette. They clearly show that the background is more noisy using threshold 95 but the reconstruction result would be better in the last section.

2 VOLUME OF INTEREST

This is another important part for later reconstruction as you need to put the whole statue inside the designed bounding box. The value of x, y, z is kind of hard to find immediately. I have tried many values and compare the result of the final output and finally my bounding box is like $[0.3, -1, -2; 2.3, 2, 2.6]$, which basically have a good result on my shape from silhouette. Here shows to angle for the rotating statue with bounding box vertices. Basically it is not obvious to judge the bounding box by these two views, but we can see the reconstruction result later.

Figure 1.1: Extracted Silhouette with threshold 95.

Figure 1.2: Extracted Silhouette with threshold 120.

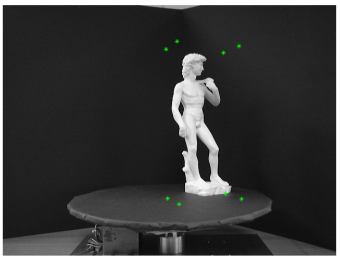

Figure 2.1: The bounding box from the front view.

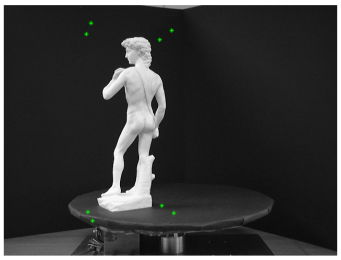

Figure 2.2: The bounding box from the back view.

3 VISUAL HULL

This is the main part for calculating the values of voxels (volume pixels) from those extracted silhouette from 18 cameras. If we project the voxel into a specific silhouette (whose corresponding pixel is foreground), we add 1 to this voxel and consider it to be a voxel in the visual hull, which is determined by the intersection of 18 silhouette cones. After going through all the voxel inside the bounding box, we can find the value of each voxel and then compute the reconstructed 3D object by iso-surface in Matlab. Assume that I fix the bounding box and try the influence of different silhouette threshold, I found the different result in Figure 3.1-3.3. Obviously, with higher threshold, the reconstruction of the arms becomes worse and therefore I found that the threshold 95 is a better choice. The fourth one is the reconstruction with lower resolution ($10 \times 10 \times 20$), which means less voxels in the bounding box. The result is much worse than other 3. And that last one show that all parts of the statue can be clearly reconstructed by a well-designed bounding box.

4 IMPROVEMENTS

During the exercise, I have some ideas to improve the model of shape from silhouette. In this task, we just use the threshold to determine the area of statue from the background and get the silhouette. But here is the problem. The environment here is quite easy to tell the foreground and background using simple threshold technique, but what about more complex environment ?

My idea is that we can automatically extracted the statue from a more complex environment with more colors, shadings, etc using depth image. I have ran the example of stereo vision problem given two neighbouring views of the same image and reconstructed the disparity map using searching on epipolar lines or Markov Random Field (MRF). After getting the depth information from the image, we can easily determine the foreground of background and filter out some noise with additional depth information. Also, we can store the depth image for those silhouettes without depth information in order to compensate the lost information in silhouette to improve the reconstruction performance using those depth information.

Another improvement I think is far more difficult is that how can we localize the statue in 3D space automatically, which seems not that easy in even 2D space. But basically there have been some works on 3D detection. For example, there is a 3D Object Detection dataset in the KITTI datasets and there are many related methods using deep learning. And I think those methods with multiple views of the same object (e.g., statue) will perform better.

These two ideas are all about automatic extraction without setting the threshold or bounding box manually. But the most important thing is for shape from silhouette I think is still the lost depth information of silhouette and what we need to improve is that we can use stereo image pair to get additional depth information iteratively with 18 cameras, without just using this simple threshold method.

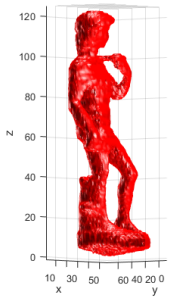

Figure 3.1: Reconstruction Object with Silhouette Threshold 95, with higher resolution ($64 \times 64 \times 128$).

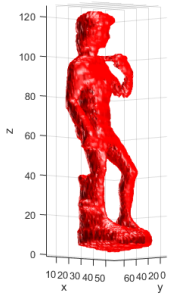

Figure 3.2: Reconstruction Object with Silhouette Threshold 105.

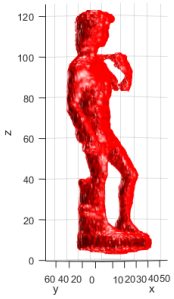

Figure 3.3: Reconstruction Object with Silhouette Threshold 120.

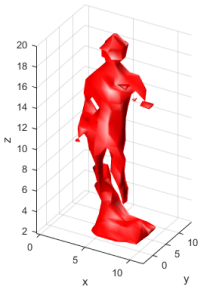

Figure 3.4: Coarse Reconstruction with lower resolution.

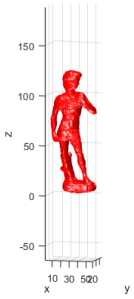

Figure 3.5: Front view by a well-designed bounding box.

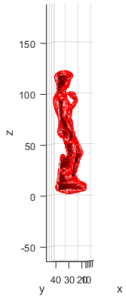

Figure 3.6: Side view by a well-designed bounding box.