Computer Vision

Assignment 9 : Shape from X

Autumn 2018

Nikita Rudin

December 6, 2018

In this lab we are implementing a stereo reconstruction from images of an object on a turntable. We are given 18 images of the objetcs the corresponding projection matrices.

1 Silhouette extraction

We start by extracting the silhouettes of the object from the images. We use a simple threshold since we have a white object on a black background. We tune this threshold manual by selecting the best visual result.
This gives us masks which we will use in to compute the visual hull.

Figure 1: Silhouette extracted from the 1st image

2 Visual hull

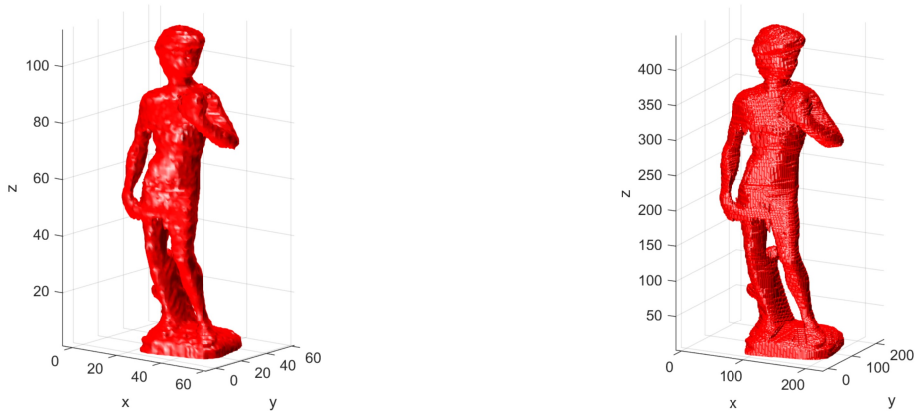
2.1 Bounding box

We will now reconstruct the 3d shape of our object. In order to simplify the problem we define a bounding box around our object. Once again this is done manually by projecting

the corner of the box into the images and trying to fit it around the object.

2.2 Voxel projection

We then divide the volume of interest into a grid of voxels (3d pixels). We project each voxel into each image and check it it lands inside the silhouette. If it does we increase the value of this voxel. Finally a provided code extracts an iso-surface for the voxels. The computation is implemented by first getting all 3d points into a long 4xN matrix. They are then projected and checked all at the same time. This replaces 4 nested for loops (x, y, z, cameras) by just the one for cameras. It runs approximately 5x faster for the 64x64x128 resolution.



(a) low resolution

(b) image 3

Figure 2: Comparison of the obtained iso-surfaces using different resolution. While the low-res version (64x64x128) is computed in 1.5s the high-res (256x256x512) took 8 minutes.

We can see that there is not a dramatic increase in quality between the two results and for most practical application the lower resolution should be used. Other than that, we can see that the results are reasonably good.

3 Parameters

- silhouette threshold: 110
- bounding box: [0.2 -0.2 -1.8; 2.5 1.5 3]
- resolution: from 64x64x128
- volume threshold: 17

4 Discussion

Even though it seems that this approach is very effective there are some drawbacks. First of all, it requires a precise set-up and calibration. We need to have a calibrated camera, know exactly how much the object turned between each frame and easily be able to separate the object from the background.
Another drawback is implied in the fact that we use silhouettes. We are not getting any information about the shape in the axis of the camera. This is mostly solved in our case by using images from all angles, but we can see that if we were to fewer cameras the results are really not great.

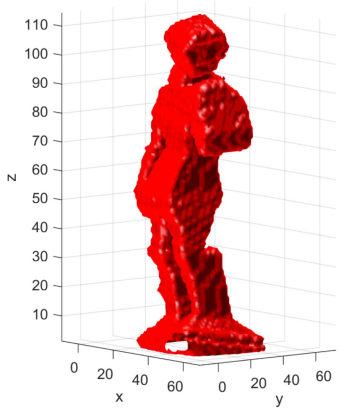


Figure 3: Computed shape by using only 2 cameras with perpendicular angles of view

Furthermore, even when using 18 cameras we see some filled areas which should actually remain empty this due to the fact that there are still points where all of the cameras are obstructed and we therefore don't get enough information about them. For example we are unable to get a good distinction between the right leg and the object behind it.
We can try to use additional information provided by the images to improve our results. We can leverage the shape from shading techniques to get better representation of cavities for example.