Advanced Robotic Perception: Lecture 9

Lecture Notes

Victor Risager

October 10, 2023

1 Mask R-CNN

Pose estimation of people.
We represent the human with x joints, where we want to know the position of each of these joints.
Done by segmenting different parts of the human, and then having preset constraints towards the different points.
Self occlusion can occur where e.g. an arm of the human is in front of the rest of the body.

1.1 Learning the segmentation

First some annotation is done.

1.2 Deep learning

Difference between handcrafted models, Machine learning, and deep learning.

- Handcrafted features are made by humans.
- Machine learning stil uses handcrafted features and learns the classification.
- Deep learning learns the features, learns the classification.

2 Mask R-CNN

It does everything. It kan also just use keypoints. It outputs an image with multiple channels, where each channel has its own class. Works well in populated and occluted scenes.

3 Pose estimation 3D

To get ground truth, you would need a massive setup with cameras. You can have a library with multiple poses, and then fit them onto images. Furthermore it is possible to use 2D images, to estimate position in 3D. Work by stefan, (Underviser).

4 Pointclouds

Using 2d monocular image can provide depth, however it can only provide relative depth, and not absolute depth in e.g. meters.
There are different 3D representation:

- Point Cloud
- Voxel Grid
- Mesh
- Multiview/Depth map. (More oldschool) (it uses gradients, which displays the depth of the image using the grayscale.)

4.1 Disparity

Using Disparity, you loose resolution the further you go away, the disparity to distance map almost looks like an inverse logarithm or $\frac{1}{x}$ where x is the distance.

Disparity to depth
Closed form solution. Can also be used for 3d using the reprojection equation, which uses the intrinsic and extrinsic camera parameters.

4.2 PCL (Pointcloud Library)

Important to use a filter. You can use the 50 nearest points to figure out the standard deviation and then if it is further away than some threshold, then it is probably an outlier.

5 Deeplearning on Pointclouds

images are always

- Grid structure
- Spatial cohesion (you know where every pixel is in the image.)

Pointclouds however

- Permutation: The order of rows does not matter. This entails that you need an algorithm which can handle all orders of rows of the point cloud.[1]

5.1 Point net

First you do a point projection which projects 3 dimensions into e.g. 64 projection, and then it does max pooling on the point features. The pooling makes it invariant to the roworder. You do max-pooling across each dimension. This works very well.
You need the correct number of input points. Therefore you can either throw away points, or dublicate points to add them in order to get e.g. 1024 points.

6 Graph neural networks

You can get neighborhood information and use cnn's to use the neighborhood information for segmentation.

7 Exam

draw 2-3 topics and then talk about the miniproject.

[1] A row is a point where the colums hold e.g. X, Y and Z.