Yihao Qian

Team A: Aware

Teammates: Amit Agarwal Harry Golash

Menghan Zhang Zihao (Theo) Zhang

ILR05

Nov.23, 2016

Individual Progress

During the last week, I was in charge of object detection. I was also in charge of stereo vision.

Object detection

The algorithm we are going to use is called Single Shot MultiBox Detector (SSD). The input of this algorithm is the image we captured from the camera. The output of this algorithm are as follow: 1. The bounding box of the interest object, which describe the relate position of the object. 2. The category that the object belongs to 3. The possibility that the object belongs to this category.

Single Shot MuliBox Detector (SSD) is a convolutional neural network. The structure of this network are as follows:

1. VGG-16-Net. This is the common convolutional neural network that is used to extract the features. The convolutional neural network is the combination of the convolutional layers and pooling layers. The convolutional layers are used to extract the features. The pooling layers are used to decrease the size of the image. For example, the input of the pooling layer is a matrix [3,4;6,7], this is a 2*2 matrix, if we use the max pooling strategy, the output of the pooling layer is 7, since 7 is the biggest number in the matrix. The pooling layer is quite useful. It can present the special relationship between the features extracted from the convolutional layers, and change the features to a high level features. VGG-16-Net is a pre-trained network. It has been proved using the VGG-16-Net can extract the features
2. Unlike Faster R-CNN, which use the Region Proposal layer in order to generate the region proposal. SSD use a series of decreasing convolutional layer and regression boxes in order to generate the region proposal. You may find the structure of the SSD in Figure 1.
3. FC (Fully connected) layer. Fully connected layer takes both the input from the RPN and the feature map, outputs the bounding box, category and the possibility.
4. The last layer is called the Non-Maximum Suppression layer. The input of this layer is all the bounding box and the corresponding labels. However, for one object (human), there may be several bounding box around it, which makes the picture very untidy. Non-Maximum Suppression layer (NMS) is used to compute the most appropriate size for the object.

We tested the performance of the SSD on my computer. My computer has a I5-4600U processor, 8G RAM, GT740M graphic card. The fps of the SSD is 2fps, which is 40 times faster than the Faster R-CNN (20s/image). With high performance GPU, such as TitanX, it would take 50-60 fps.

Camera calibration:

Test set up:

Fixed the stereo vision in the car. Make sure the relevant position of those two cameras are fixed. Put the chessboard in front of the stereo camera. Catch the image from stereo camera simultaneously. Change the position and rotation of the chessboard.

After receiving 20-30 pairs of images from stereo camera, Matlab could compute the intrinsic and extrinsic parameter of the camera. Matlab could also compute the relevant translation and rotation from camera 2 to camera 1. With those parameters we could calculate the depth information. You may find the depth information of the chessboard in the Figure 2.

Challenges:

Calculating the depth:

The parameters that Matlab provides to us is very different from what we learnt in CV classes. For example, the fundamental matrix & intrinsic & extrinsic matrix is the transposed matrix of the true value. After searching on line about the meaning of this, I could finally figure out how to use it.

Teamwork:

I read the paper for object detection (SSD). I found out SSD is the best algorithm for our project, from accuracy and fps perspective. Zihao and Amit tried to let the radar work. Harry was in charge of designing the rack of the camera.

Plans:

From now on, we are going to build our own perception system.

Harry and Amit will finish the camera housing design and test the performance of the housing.

Zihao and Amit are going to test the data from the radar.

I am going to finish the stereo vision.

Figures:

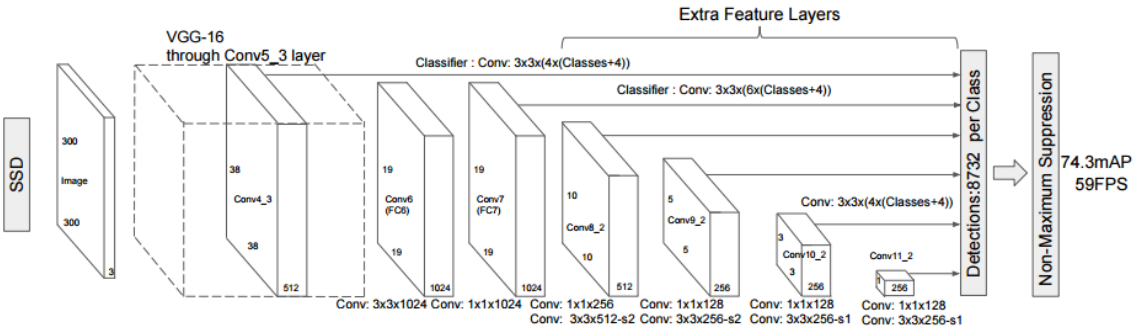

Figure 1. SSD network Structure

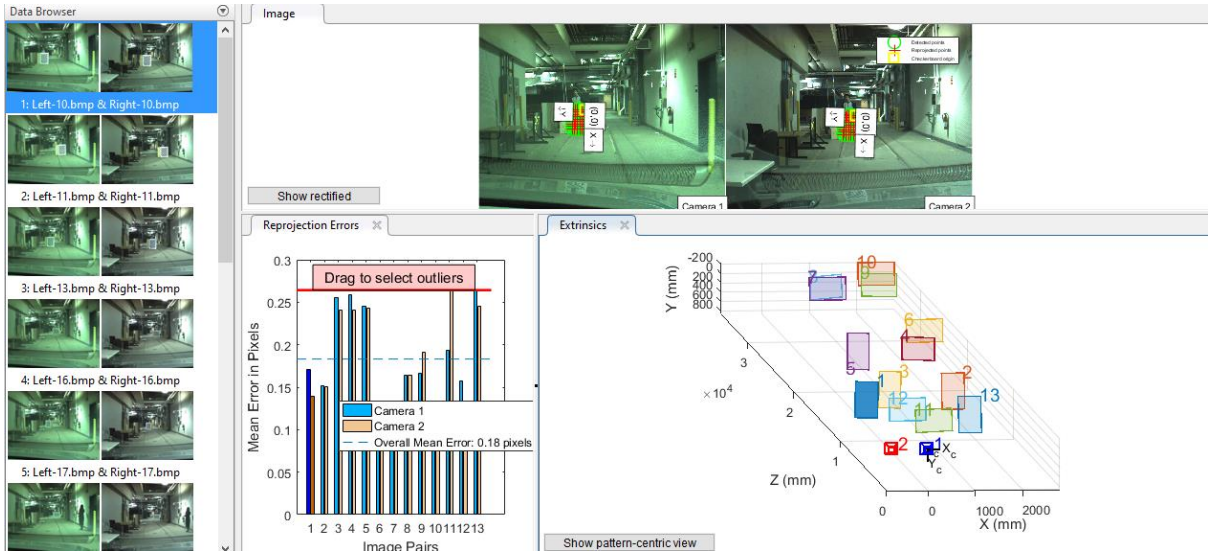

Figure 2. Stereo Vision