Contents

1 Introduction

Stereo image matching estimates the disparity between a rectified image pair, which is of great importance to depth sensing, autonomous driving, and other related tasks.

The left image features and the right image features are divided into groups along the channel dimension, and correlation maps are computed among each group to obtain multiple matching cost proposals, which are then packed into a cost volume. Group-wise correlation provides efficient representations for measuring feature similarities and will not lose too much information like full correlation. It also preserves better performance when reducing parameters.

2 Problem Definition

To estimate the disparity images for predicting the depth in an image based on the stereo(left,right) images of cameras which are separated by a fixed distance.Here we use Guided Aggregation network.

3 Objectives

- To estimate disparity maps for stereo matching, which incorporates to build up the cost volumes through GAN.
- To explore the Agrregation cost by decreasing channels in the 3D aggregation network to verify the effectiveness of GAN.
- To improve the accuracy by implementing the stacked hourglass backbone and considerably increasing the number of 3D convolutional layers for cost aggregation.

4 Requirement Analysis

- Ubuntu 16.04 and Above.
- GPU
  - Nvidia Tesla K40P and Above.
  - Nvidia Quadro P400
- Nvidia Drivers and Cuda.
  - Driver: Version 440.33.01
  - Cuda: Version 10.0 and Above.
- Python Packages
  - Pytorch : 1.2.0
  - TensorFlow : 2.0
  - Protobuf : 3.11.0
  - Torchvision :0.4.0
  - numpy : 1.15.4
- Datasets
  - KITTI 2012, KITTI 2015, Sceneflow Datasets.

5 High Level Design

- **Model**

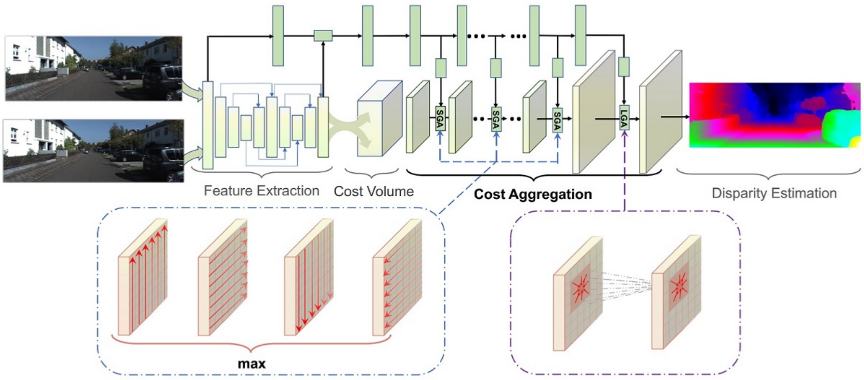

Fig 1.1 : Gudide Aggregation streo network

- Feature Extraction
- Cost Volume
- Cost Aggregation
- Disparity Estimation

- **Feature Extraction**

  For the feature extraction, we use a stacked hourglass network which is densely connected by concatenations between different layers. The feature extraction block is shared by both left and right views. The extracted features for left and right images are then used to form a 4D cost volume

- **Cost Volume**

  The cost volume The extracted features for left and right images are then used to form a 4D cost volume.

- **Cost Aggregation**

  Cost Aggregation consists of two part

  – Semi-Global Guided Aggregation (SGA)

  – Local Guided Aggregation (LGA)

- **Semi-Global Guided Aggregation (SGA)**

  – First, SGM has many user-defined parameters (P1,P2), which are not straightforward to tune. All of these parameters become unstable factors during neural network training.

  – Second, the cost aggregations and penalties in SGM are fixed for all pixels, regions and images without adaptation to different conditions.

  – Third, the hard-minimum selection leads to a lot of fronto parallel surfaces in depth estimations. We design a new semi-global cost aggregation step which supports backpropagation. This is more effective than the traditional SGM and can be used repetitively in a deep neural network model to boost the cost aggregation effects.

  – The proposed aggregation step is:

$$C_r^A(p,d) = w_0(p,r).C(p,d) + w_1(p,r).C_r^A(p-r,d)$$

$$+w_2(p,r).C_r^A(p-r,d-1) + w_3(p,r).C_r^A(p-r,d+1)$$

$$+w_4(p,r).max_i C_r^A(p-r,i)$$

- **Local Guided Aggregation (LGA**

  - The LGA layer learns several guided filters to refine the matching cost and aid in the recovery of thin structure information.
  - The local aggregation follows the cost filter definition.

- **Disparity Prediction**

  For the disparity estimation, we employ the disparity regression proposed in:

$$\hat{d} = \sum_{d=0}^{Dmax} d(-C^A(d))$$

  The disparity prediction dˆ is the sum of each disparity candidate weighted by its probability. The probability of each disparity d is calculated after cost aggregation via the softmax operation (·). The disparity regression is shown more robust than classification based methods and can generate sub-pixel accuracy.

6 Conclusion

- The GAN shows much more efficient and effective guided matching cost aggregation (GA) strategies, including the semi-global aggregation (SGA) and the local guided aggregation (LGA) layers for end-to-end stereo matching.
- To Show that significantly improve the accuracy of the disparity estimation in challenging regions, such as occlusions, large textureless/reflective regions and thin structures.
- To replace computationally costly 3D convolutions and get better accuracy