

CFM AI&VR Weekly Report: Week #15 of 2019

Zhang Yifei(yidadaa@qq.com)

UESTC — April 11, 2019

1 Summary

Recently, I read lots of literature about how to make Depth Prediction from mono video sequences and I will introduce two of them next. At the same time, I am preparing for two competitions.

2 Details

2.1 Paper Reading: Unsupervised Learning of Depth and Ego-Motion from Monocular Video Using 3D Geometric Constraints

The paper present a novel approach for unsupervised learning of depth and ego-motion from monocular video, which could be intergated to Visual Slam System. The main contribution is to explicitly consider the inferred 3D geometry of the whole scene, and enforce consistency of the estimated 3D point clouds and ego-motion across consecutive frames.

2.2 Paper Reading: Depth Prediction Without the Sensors: Leveraging Structure for Unsupervised Learning from Monocular Videos

The paper proposes a novel approach which produces high quality results of depth prediction, which is able to model moving objects and is shown to transfer across data domains. The main idea is to introduce geometric structure in the learning process, by modeling the scene and individual objects(according to semantic segmantation produced by MaskRCNN). Camera ego-motion and object motions are learned from monocular videos as input.

2.3 Competition: OPPO Top AI Competition

The final stage of the competition will be hosted at **Friday of next week**. Out team is focusing on making demos and preparing presentation materials. We are planning to write an app to demonstrate out results. However, we found that the result of our network doesn't seem to work well on real-world image and we are trying to improve it.

2.4 Competition: Cainiao Global Technology Competition 2019

The first stage of the competition will come to an end at next **Monday**. Competitors are required to give a solution about how to measure the size of an object on a smart phone. We determine to use VINS Mono SLAM system to solve the problem. First, the 3D point clouds are generated by VINS Mono SLAM; Second, we use DeepLab V3 to get the semantic information of every frame recorded by phone, and get the target points from point clouds; Finally, we calculate the size of the object according to the points.

3 Plan

- Go on with reading papers about Depth Prediction, I am planning to publish a paper about it.
- Try my best in the competitions.