CS 766 - Project Proposal

Training A Deep Network for Multi-view Disparity Estimation

Xucheng Wan	Zhengyang Lou	Xiao Wang
9077946623	9073000235	9078040434
xucheng.wan@wisc.edu	zlou4@wisc.edu	xwang2229@wisc.edu

Problem Statement:

In this project, we will conduct research on applying deep learning in multi-view disparity estimation. Disparity estimation is essential to many challenging tasks in computer vision like 3D reconstruction and 3D object tracking[1].

In the setting of binocular disparity, where two cameras are used, disparity is the value that characterizes the displacement of two pixels from left and right images representing the same point in the scene. However, some problems may occur when having only two views. For example, occlusion is the problem when some objects are only captured by one camera, in which case no stereo information is available.

To solve these problems, multi-view disparity estimation algorithms are proposed. In multi-view disparity estimation, multiple pictures of the same scene from different viewpoints are used to estimate the disparity value, which is shown in the following figure. Even though multi-view stereo is a natural extension of binocular, multi-view stereo gradually became a unique problem itself, for instance, multi-view stereo algorithm may take a set of images surround an object as input [2]. For this project, we will only consider the first step of multi-view stereo, namely multi-view disparity estimation.

Current Research:

(1) Binocular Disparity Estimation using Deep Learning

Most research using deep learning on disparity estimation focuses on binocular disparity estimation. We briefly summarize three recently works to illustrate the current state-of-art. In 2015, Zbontar and LeCun [3] first proposed to use neural networks to solve the problem of disparity estimation. Their algorithm uses a sigmanese network to compute matching cost, then performs post-processing to get the disparity map. Zbontar and LeCun's algorithm was the best performance algorithm when their paper was first published. Later, Luo, Schwing, and Urtasun [4] introduce a faster and more accurate network, which replaced the concatenation of left and right patches with a dot product layer to speed up the process. Luo et all's algorithm also learns the probability distribution of all possible disparity values rather than the a matching score. Then in ICCV 2017, Kendall, Matirosyan, Dasgupta, and Henry [5] proposed a new network structure called Geometry and Context Network (GC-Net) which uses fully connected convolutional neural network to compute the disparity map end-to-end. Amazingly, the GC-Net improves the state-of-art in KITTI 2015 and KITTI 2012 (two important data sets) by about 10% and 20%. And this GC-Net is also faster than most of the other methods.

(2) Multi-view Disparity Estimation

Even though having more views helps to ease a lot of problems encountered by binocular disparity estimation, it may not be optimal to use all views in multi-view disparity estimation. Zheng, Dunn, Jojic, and Frahm [6] propose a pixel-level jointly view selection and depth estimation algorithm based on Hidden Markov Random Field.

Proposed Method:

In this project, we will investigate the potential of deep learning in multi-view disparity estimation. Our goal is to develop an deep learning solution for multi-view disparity estimation. We will test two possible solutions. The first one is to first perform binocular disparity estimation using an existing deep learning method for every image pairs in the set of images. Then perform fusion to come up a single disparity map based. The second approach will use a fully convolutional neural network to determine the views to use for each pixel location. After the selection of views, the algorithm will then use an existing multi-view disparity estimation method to compute the disparity.

We believe our method will provide a better disparity map for the following two reasons. It has more views than deep learning based binocular disparity estimation algorithms. And traditional multi-view disparity estimation algorithms have a lot of hand-crafted features which may not be the optimal choice while neural network may provide better feature representation. So we believe our method will have a better performance.

Plans and Evaluation:

Our project consists of three main steps:

Step 1: Implement two methods in Python- one traditional method on multi-view disparity estimation and one deep learning based binocular disparity estimation. We will refer to Ref. [4] and Ref. [7] for guidance. If time permitting, we will try to implement [5] and [6].

Step 2: Investigate the availability of using deep learning on multi-view disparity estimation and conduct experiments on the proposed algorithm.

Step 3: Compare results from our new method with other methods, including multi-view and bomoular algorithms. We will use dataset Fountain-P11 and Herzjesu-P9 [8]. The evaluation criteria would be the percentage of pixels having the same or similar disparity value as the ground truth. Besides, zhou et all's method [7] deals with disparity estimation in noisy condition. So we will also compare our method with their method in noisy condition.

References:

[1] Mattoccia. Stereo vision: algorithms and applications.

[2] Y. Furukawa and C. Hernández. Multi-View Stereo: A Tutorial.

[3] J. Zbontar and Y. LeCun. Computing the stereo matching cost with a convolutional neural network. CoRR, abs/1409.4326, 2014.

[4] W. Luo, A. Schwing, and R. Urtasun. Efficient deep learning for stereo matching. In CVPR, 2016.

[5] A. Kendall, H. Martirosyan, S. Dasgupta, P. Henry, R. Kennedy, A. Bachrach, and A. Bry.

End-to-end learning of geometry and context for deep stereo regression. arXiv preprint arXiv:1703.04309, 2017.

[6] E. Zheng, E. Dunn, V. Jojic, and J.-M. Frahm. PatchMatch based joint view selection and depthmap estimation. CVPR 2014

[7] S. Zhou, Z. Lou, Y. Hu and H. Jiang. Improving disparity map estimation for multi-view noisy images. International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2018.

[8] C. Strecha, W. Hansen and L. V. Gool. Combined depth and outlier estimation in multi-view stereo. In CVPR, 2006.