Research Proposal - Plan

Student: Davit Buniatyan
ID: 13008791

Proposed Research Topic: Depth map reconstruction from sequential images using Recurrent Neural Networks.

Purpose & Background: There has been extensive research in Computer Vision for generating depth map from single or more images. State-of-the-art technique is considered to be Make3D, which uses Markov Random Field to infer set of "plane parameters" that capture both the 3D location and 3D orientation of the patch (2009). K. Konda & R. Memisevic suggested unsupervised learning method using autoencoder to extract depth and motion trained on set of images either varied in time, in location (different cameras) or both (2013). With activation of deep learning methods in computer vision, a multi-scale deep network has been proposed to predict depth map using single image (2014). Convolutional Neural Network has been trained on Stereo Images with the error ~2.61% on KITTI dataset. Recently, Google researchers published paper called DeepStereo, which is based on dual convolutional neural network architecture that also reconstructs depth map for directly generating new views from imagery (2015). In the avoidance of limited input space, Recurrent Neural Networks (RNN) are considered to be very effective machine learning technique to be applied on sequential data such as sequential text, images (or video), signals, etc.. The purpose of this research is to architect RNN for generating depth map from sequential images and benchmark on state-of-the-art techniques and other deep learning methods the computational complexity, runtime complexity & error rate. The results potentially could be used in robotics for 3D reconstruction of the rigid and non-rigid dynamic objects and environments or in Visual FX post-editing.

Scope: The following research is going to be my 3rd year project. I am planning to conduct literature review, collect data and draft architecture during my first term and spend my full second term on training and refining the RNN. The method of data generation may vary such that it might be possible to conduct real-life experiments (extraction of ground truth depth from video games) or scrape from Internet resources (e.g. Youtube Stereo videos).

Theoretical framework: Given the need to synthesize data, it will be collected with the reference to generating Flying-Chairs in FlowNet paper (2015). I will use deep learning framework most probably Theano (others Caffe with RNN & LSTM plugin or Torch7) for constructing RNN and training the model. During data collection, testing set will be separated to avoid overfitting the model. The trained algorithm will be also be tested on KITTI and NYU Depth dataset to benchmark results.

Method:

1. Conduct Data Collection
   As current datasets are not enough for training RNN, I propose three methods for data collection.
   - Scraping Stereo Videos from Youtube and reconstructing depth map via computing stereo matching costs with CNNs described in LeCun's paper (2015): The advantage of this method is that we have vast amount of available resources, however given the error of ˜2.61% our results will be artificially lowered minimally by ˜2.61%.
   - Capturing ground truth depth map and RGB while playing high graphics video games (such as GTA 5, Far Cry 3, Battlefield 4): The advantage of this method that we will have vast amount of ready to use simulated environments for training that will supposably generalize, (e.g. generalization of FlowNet from Sintel Dataset), however capturing the depth map will need additional activity and time to spend on playing games.
   - Rendering custom 3D images using either game engine (Unity3D for fast and programmable results) or 3Ds Max with Vray renderer (for realistic results). This might take time and computational resources for generating high quality dataset, however depth variations of the same images will be feasible to generate in order to significantly enhance training.
2. Recurrent Neural Network
   a. Developing simple RNN and training on simple data in order to learn how to effectively construct the model via chosen framework.
   b. Architecture prototype of the RNN.
   c. Running Feedback loop of training the neural network and adjusting parameters.
3. Analyse results
   a. Run results on KITTI and NYU Depth datasets and benchmark with state-of-the-art techniques
   b. Write research article & report

Plan:

Prepare proposal by	4 October
Complete literature review by	1 November
Conduct Data Collection by	1 December
Prototype RNN architecture by	15 December
Train and fine-tuning RNN by	15 March
Analyse results by	1 April
Complete final article & report by	16 April

Limitations: The collected data might not be good for getting high results. The training data might not generalize well such that the architecture of RNN will be data specific. The computational and runtime complexity might stand against generating fast results.