Technical Review Paper Evaluation Form

(attach this form as the cover page for your report)

Student Name: Yifan Shen

Project Advisor: Fuming Zhang

Team Name: XXLS

Team Members: Fanzhe Lyu, Ruoyang Xu, Yilun Xie

_______ / 30 Technical Content

- Current state-of-the-art and commercial products
- Underlying technology
- Implementation of the technology
- Overall quality of the technical summary

_______ / 30 Use of Technical Reference Sources

- Appropriate number of sources (at least six)
- Sufficient number of source types (at least four)
- Quality of the sources
- Appropriate citations in body of text
- Reference list in proper format

_______ / 40 Effectiveness of Writing, Organization, and Development of Content

- Introductory paragraph
- Clear flow of information
- Organization
- Grammar, spelling, punctuation
- Style, readability, audience appropriateness, conformance to standards

_______ / 100 Total - Technical Review Paper

Deep Neural Network-based Visual Odometry Estimator for Indoor Blimps

Introduction

In robotics and computer vision, visual odometry is the process of determining the position and orientation of a robot by analyzing the associated camera images and has been widely utilized in a variety of applications. Use cases for visual odometry (VO) include motion estimation for vehicle and point-cloud mapping. Unlike traditional approaches that rely on monocular video or use stereo information, the recent learning based hybrid VO approaches that capitalize on convolutional networks are proved to be effective in tasks like classification, localization and depth estimation [1]. This technical review briefly summarizes some of the state-of-the-art deep learning-based visual odometry estimator, introduces underlying neural network computation mechanisms in computer vision, and provides an ideal implementation of neural network-based visual odometry estimator for indoor environment.

Commercial Application of Visual Odometry Estimator

Most of existing VO algorithms are developed under a standard pipeline includes feature detection, feature matching, motion estimation and local optimization [2]. While hybrid visual odometry estimation that utilized end-to-end, deep learning-based architecture recently shows promising results in various applications. From the sequential image inputs, representations of depth and motion are extracted by detecting synchrony across time and stereo channels using network layers with multiplicative interactions. The extracted representations are turned into information about changes in velocity and direction using a convolutional neural network (CNN) [1].

The Deep Neural Network (DNN) system from Nvidia learns a map representation captured by camera and outputs an estimation of the camera location within the environment [3]. The map representation may be updated through supervised training and /or self-unsupervised training. Additionally, the DNN can be trained by fusing multi-sensory inputs with the camera poses estimated by the DNN to improve accuracy.

TuSimple's neural network architecture system for deep odometry combines DNN with merge modules in the visual odometry model [4]. The Convolutional Neural Network (CNN) extracts representative features from input images, merges the outputs and generates a flow output for layers of DNN. The DNN then generates flow output for prediction of static optical flow and the motion parameters.

Since the development of existing VO estimator mainly depends on algorithm improvement, the cost of

VO system is hard to estimate from the sources discussed in this review paper.

Underlying Mechanisms: DNN/CNN Architecture and End-to-end Framework

Convolutional Neural Networks consist of two major components: convolution layers, which perform matrix convolution operations and apply several filters to the target input to detect specific features in the input, and fully connected layers, which process the resulting detected features and multiply the data matrices with weight matrices to determine the class of the input [5]. Thus, computing the output of the neural network requires two major operations: matrix convolution and matrix multiplication.

The measures of a hybrid VO system performance include accuracy. Since the VO estimator is built on a blimp platform, power and speed are also under consideration. Modern CNN architectures include both deep and shallow ones. Deep CNN architectures, like Vgg-16 or Vgg-19, use tens of convolutional layers followed by several layers of fully connected neurons [7]. Shallow CNN architectures, like LeNet-5, use only two to three convolutional layers followed by one to two layers of fully connected neurons [5]. Deeper CNNs tend to have 5%-30% higher accuracy but takes more computation power in orders of magnitudes to train and to compute inference than shallower ones [7].

End-to-end framework of VO system means the DNN and/or CNN infers poses directly from raw RGB images without adopting any modules in the conventional VO pipeline [6]. The end-to-end system mainly relies on composed of convolutional neural network based feature extraction and recurrent neural network based sequential modeling.

Building Blocks of an Ideal Deep Learning-Based Visual Odometry Estimator for Indoor Blimps

To estimate the state of indoor blimps from visual odometry estimator, both software algorithms and hardware computations are required. Software algorithms are used to extract features from input images of the indoor environment, to estimate the motion and orientation of the blimp and to perform local optimization for higher accuracy. The hardware system includes an Internet server, comprising I/O port that configured to receive signals from the client devices (blimps), a memory and several or more processing units for the parallel computation of CNN. Also the blimp needs to be equipped with a rotatable camera or sensors for environmental image capture and an adequate power supplier for the continuous work of the camera. The communication between the platform and the blimp can be through Bluetooth or Wi-Fi.

References:

[1] Konda, Kishore and Memisevic, Roland, “Learning Visual Odometry with a Convolutional Network,” *in Proceedings of the 10th International Conference on Computer Vision Theory and Applications*, 2015.

[2] Patrick McGarey, *CSC2541 Visual Perception for Autonomous Driving*. [Online]. Available: http://www.cs.toronto.edu/~urtasun/courses/CSC2541/03_odometry.pdf. [Accessed: 28-Sept-2019].

[3] NVIDIA Corporation, “Learning-Based Camera Pose Estimation From Images Of An Environment,” U.S. Patent, No. 20190108651, April 11, 2019.

[4] TUSIMPLE (San Diego, CA, US), Zhu, Wentao, Wang, Yi and Luo, Yi, “NEURAL NETWORK ARCHITECTURE SYSTEM FOR DEEP ODOMETRY ASSISTED BY STATIC SCENE OPTICAL FLOW,” U.S. Patent No. 20190079534, March 14, 2019, Available: http://www.freepatentsonline.com/y2019/0079534.html. [Accessed: 29-Sept-2019].

[5] Y. LeCun, *MNIST Demos on Yann LeCun's website*. [Online]. Available: http://yann.lecun.com/exdb/lenet/. [Accessed: 29-Sept-2019].

[6] Sen Wang, Ronald Clark, Hongkai Wen and Niki Trigoni, “DeepVO: Towards End-to-End Visual Odometry with Deep Recurrent Convolutional Neural Networks,” in *2017 IEEE International Conference on Robotics and Automation (ICRA 2017)* [Online]. Available: https://arxiv.org/abs/1709.08429v1. [Accessed: 29-Sept-2019].

[7] Simonyan, Karen, Zisserman, and Andrew, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” *[1409.1556] Very Deep Convolutional Networks for Large-Scale Image Recognition*, 10-Apr-2015. [Online]. Available: https://arxiv.org/abs/1409.1556. [Accessed: 29-Sept-2019].