SLAM (Simultaneous Localization and Mapping)

Shaif Chowdhury

April 17, 2020

Project Summary

This proposal introduces a unique approach to VISUAL SLAM for Autonomous Robotic Exploration. It nicely combines various techniques from deep learning and multi view geometry to provide a robust solution to the SLAM problem.Our system has the following contributions :

Pose Estimation : We propose to use Convolutional neural Network based approach to improve the camera localization[13] accuracy.

Loop closure Detection : Loop closure [19] is actually an Image Retrieval problem. We propose to use Unsupervised hashing to address this problem. There have been significant improvements to Hashing techniques due to use of Deep learning. Using hashing would significantly lower space and search time while giving real-time solution.

Optimization : There are different algorithms that can do this task: Kalman filtering, Particle filtering, and Bundle Adjustment [4]. We would concentrate on deep learning based bundle Adjustment.This has to be evaluated on a large scale data-set to establish the success of deep learning based Bundle Adjustment.

Intellectual Merit:

The proposed methods would make multiple novel contributions to Computer Vision and AI in general. We aim to provide novel solutions to the problem of Unsupervised hashing which would be very useful in various domains of Vision,.NLP etc. Other than that it contributes to the research in creating discrete features from large scale data. Other than that our work in Bundle Adjustment would contribute to the challenging Structure from Motion problem.

Broader Impacts:

SLAM is used in various applications like self-driving cars, unmanned aerial vehicles, autonomous underwater vehicles, planetary rovers, newer domestic robots and even inside the human body. Solution to the SLAM problem is viewed as a ‘holy grail’ for the mobile robotics community since it can make robots truly autonomous.

Our solution intends to improve the state of the art when it comes to autonomous vehicle navigation or localizing robots. The work described will also lead to academic papers in peer-reviewed journals, leading to more research in the related fields.

PROJECT DESCRIPTION

1 Introduction

There has been some success in development of SLAM[6] which has attracted the attention of big name tech companies. However, there is a lot of research on how to unify the interface of existing or emerging algorithms, and effectively perform benchmarks about the speed, robustness and portability. In this paper, we propose a novel SLAM platform that effectively uses deep learning and multi view geometry to get much better performance. The core contribution is the use of Unsupervised hashing for loop closure detection and creating a deep learning based bundle Adjustment. Our system would be suitable for a lightweight localization, leveraging visual odometry tracks for Bundle Adjustment that allow for zero-drift localization.

The remainder of this description will examine current research (Section 2). We will then outline our specific objectives (Section 3), how we intend to approach this work (Section 4), the expected outcomes (Section 5), some potential applications for this work (Section 6) and finally a suitable timeline for the project(Section 7).

2 Current research

The SLAM problem is about placing a robot at an unknown location in an unknown environment for the robot to build a consistent accurate map of this environment while simultaneously localizing itself within this map. Even though initially it appears to be a chicken-and-egg problem there are a few algorithms known to solve it, at least approximately, in finite time and resource for certain environments.

Probabilistic SLAM

Probabilistic SLAM[18] involves finding an appropriate representation for the current pose along with efficient, consistent computation of the prior and posterior distributions. The most common representations are done with additive Gaussian noise, leading to the use of the extended Kalman filter (EKF) to solve the SLAM problem.Another way is to use Recursive Bayes Filters for estimating the pose of a mobile robot. There are various implementations such as discrete filters (histograms), particle filters, or Kalman filters.

EKF-SLAM

EKF SLAM[5] makes use of an extended Kalman filter (EKF) for simultaneous localization and mapping (SLAM).EKF SLAM[12] algorithms are generally feature based, and use the maximum likelihood algorithm for data association. For a long time, EKF SLAM had been the de facto method for SLAM, until the introduction of FastSLAM.

EKF SLAM[14] is associated with gaussian noise assumption, which significantly impairs EKF SLAM's ability to deal with uncertainty. With greater uncertainty in the posterior, the linearization in the EKF fails.

Lidar SLAM

The Lidar SLAM[19] system is reliable in theory and technology. There are quite a few 2D Lidar SLAM[[20] approaches based on probabilistic methods. There are also feature based approaches from LIDAR point clouds.

VISUAL SLAM

With the development of CPU and GPU, the capability of graphics processing has become more and more powerful. With camera sensors getting cheaper, more lightweight and more versatile at the same time, the past decade has seen the rapid development of Visual SLAM[5].

The method of utilizing information from an image can be classified into direct method and feature based method. Direct method leads to semiDense and dense construction while feature based method causes sparse construction.

MonoSLAM: It is the first real-time mono SLAM[5][2] system, based on EKF.

PTAM: This one[15]) makes use of parallel tracking and mapping. It firstly adopts Bundle Adjustment to optimize and concept of key frame to build pose graphs.

ORB-SLAM: It uses[11] three different threads: Tracking, Local Mapping and Loop Closing. It supports monocular, stereo, and RGB-D cameras.

proSLAM: This is a lightweight visual SLAM system with easy understanding.

ENFP-sfm: It is a feature tracking method which can efficiently match feature point correspondences among one or multiple video sequences

OpenVSLAM: This[17] is based on an indirect SLAM algorithm with sparse features. The excellent point of OpenVSLAM is that the system supports perspective, fisheye, and equirectangular, even the camera models you design.

3 Objectives

The proposed method will be based on the following sequence of actions:

Pose Estimation : This would be done using a convolutional neural network based approach for estimating the relative pose between two cameras. The proposed network takes RGB images as input and directly produces the relative rotation and translation as output. The system would be trained in an end-to-end manner.

Loop closure Detection : Loop closure is the problem of recognizing a previously-visited location [8][7]. This is usually done by storing features from every image and searching through them. We propose to solve this by using Unsupervised hashing [8][7][3][16]. Recent studies on vision and learning have shown that hash codes can be used for processing massive amounts of images using minimized storage and computation. In particular, hashing using deep learning can be very useful for Image retrieval in real time scenarios. But, deep hashing generally needs labeled images and unsupervised hashing hasn't shown satisfactory accuracy due to either the relaxed optimization or absence of pairwise similarity. This work would focus on generating binary features from images while preserving image similarity, so that it can be used for searching or classification tasks.

Optimization : To continuously optimize the whole 3D model as well as all the camera poses. There are different algorithms that can do this task: Kalman filtering, Particle filtering, Bundle Adjustment[1]. We aim to solve this via feature based bundle adjustment (BA)[1][9], which would enforce multi-view geometry constraints in the form of feature-metric error. The whole pipeline has to be differentiable so that the network can be trained using backpropagation. The whole system nicely combines domain knowledge (i.e. hard-coded multi-view geometry constraints) and deep learning (i.e. feature learning and basis depth maps learning) to address the challenging dense SfM problem.

4 Approach

We propose to undertake the research in the following sequence:

Pose Graph

This is about building a CNN based model for calculating the relative motion of two frames [8]. The training and testing are carried out on a public dataset. The experimental results should establish superiority for known scenes.

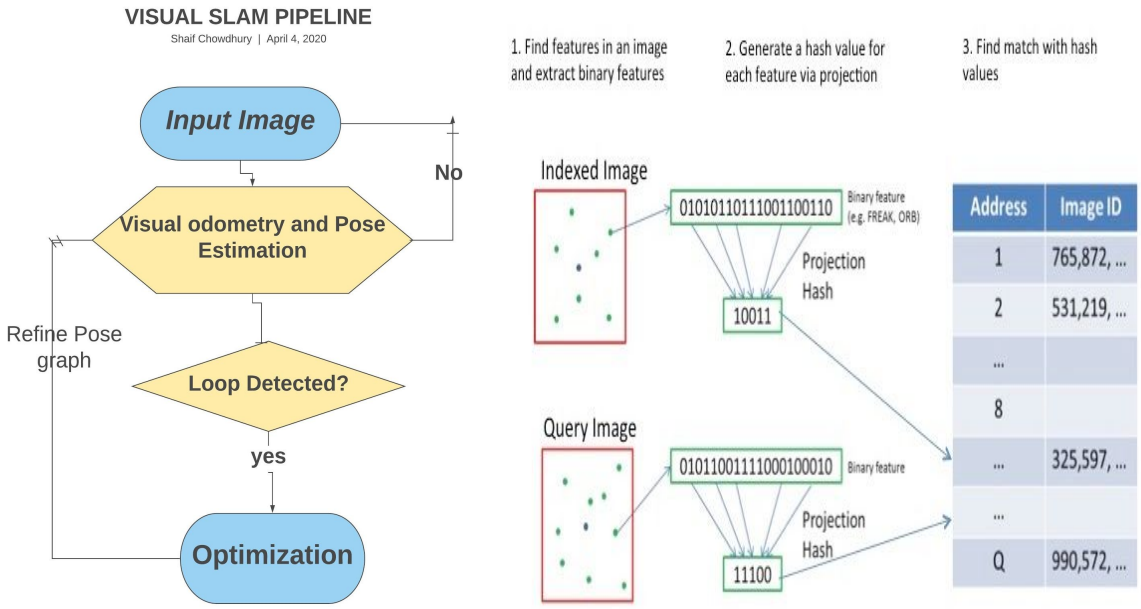
Deep Unsupervised Hashing

So, an autoencoder[14] type architecture would be needed to generate hash codes from images. Autoencoders[10] are neural networks that compress the input into a latent-space representation, and then reconstruct the output from this representation.

In our case the latent representation to binary from. For this we will need a differentiable activation function that can be used over one or multiple layers to force the latent representation to binary.

BA-Net

This is about building a network architecture to solve the structure-from-motion (SfM) problem via feature-metric bundle adjustment (BA), which explicitly enforces multi-view geometry constraints in the form of feature-metric error. The whole pipeline has to be differentiable so that the network can learn suitable features that make the BA problem more tractable. The network first generates several basis depth maps according to the input image and optimizes the final depth as a linear combination of these basis depth maps via feature-metric BA.



These diagrams show the proposed Visual SLAM Pipeline along with the use of hashing for Loop Closure detection

5 Outcomes

The proposed approach is designed to take advantage of the numerous pre existing data sets across many disciplines.

MIT Places Dataset : This is a 10 million image database for scene recognition. Because of the volume, it can be used for training the Hashing Network.

KITTY : This dataset contains images from Autonomous vehicles driving through a mid-size city with images captured by cameras and laser scanners.

The Alderley Dataset : The dataset contains images in two different conditions for the same route: one on a sunny day and one during a rainy night, making it a challenging dataset for testing loop Closure.

As final outcomes of this project, we will produce:

- One or more scientific publications in peer-review journals, along with conference papers as needed.
- A unique approach for CNN based camera pose estimation.
- A novel approach for Loop Closure using Unsupervised hashing.
- A Bundle Adjustment framework that can make use of deep learning.
- Putting it all together to create a lightweight, cross-platform SLAM system.

6 Potential applications of this research

In this section we describe some present-day and future areas where the fruits of this research would be useful.

Commercially, Visual SLAM is still in its infancy. Even though it has enormous potential in a wide range of areas, it's still an emerging technology. With that being said, it would make a huge im[act in augmented reality (AR) applications. Projecting virtual images onto the physical world requires an accurate mapping of the physical environment, and only Visual SLAM technology is capable of providing this level of accuracy.

Visual SLAM systems are also used in a wide range of field robots. For example, rovers and landers for exploring space use visual SLAM systems to navigate autonomously. Robots in agriculture, as well as drones, can use this technology to independently travel around crop fields. Autonomous vehicles can use visual SLAM systems for mapping, localization and understanding the environment around them.

Another opportunity for Visual SLAM systems is to replace GPS tracking and navigation in certain applications. GPS isn't useful indoors, or in big cities where the view of the sky is obstructed, and they're only accurate within a few meters. Visual SLAM systems solve each of these problems as they're not dependent on satellite information and they're taking accurate measurements of the physical world around them.

7 Project Timetable

Year 1 : Work on building the Architecture for Unsupervised Hashing. Test the hashing accuracy for image Retrieval tasks on public datasets. Once it shows significant improvements from available hashing techniques use it for loop closure detection on SKAM datasets. This would need the development of an approximate hash table based searching method for query images.

Year 2 : Develop a CNN based camera pose estimation that can be trained in an end to end way to generate camera pose graph. Test it on popular SLAM datasets to compare with available methods.

Year 3 : Develope Bundle Adjustment based on a deep learning architecture. Test it on KITTY dataset and compare the performance with other methods

Year 4 : Putting it all together to create a robust lightweight SLAM system. Compare it with other open source or commercial SLAM systems and work on minimizing the time and space requirements.

References cited

1. Civera, Javier, Andrew J. Davison, and José María Martínez Montiel. 2011. *Structure from Motion Using the Extended Kalman Filter*. Springer Science & Business Media.
2. Davison, Andrew J., Ian D. Reid, Nicholas D. Molton, and Olivier Stasse. 2007. "MonoSLAM: Real-Time Single Camera SLAM." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29 (6): 1052–67.
3. Deng, Cheng, Erkun Yang, Tongliang Liu, Jie Li, Wei Liu, and Dacheng Tao. 2019. "Unsupervised Semantic-Preserving Adversarial Hashing for Image Search." *IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society* 28 (8): 4032–44.
4. Di, Kaichang, Qiang Zhao, Wenhui Wan, Yexin Wang, and Yunjun Gao. 2016. "RGB-D SLAM Based on Extended Bundle Adjustment with 2D and 3D

Information." *Sensors* 16 (8). https://doi.org/10.3390/s16081285.
5. Esparza-Jiménez, Jorge Othón, Michel Devy, and José L. Gordillo. 2016. "Visual EKF-SLAM from Heterogeneous Landmarks." *Sensors* 16 (4). https://doi.org/10.3390/s16040489.
6. Fernández-Madrigal, and Juan-Antonio. 2012. *Simultaneous Localization and Mapping for Mobile Robots: Introduction and Methods: Introduction and Methods*. IGI Global.
7. Garcia-Fidalgo, Emilio, and Alberto Ortiz. 2018. *Methods for Appearance-Based Loop Closure Detection: Applications to Topological Mapping and Image Mosaicking*. Springer.
8. Guo, Fei, Yifeng He, and Ling Guan. 2017. "RGB-D Camera Pose Estimation Using Deep Neural Network." *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. https://doi.org/10.1109/globalsip.2017.8308674.
9. Im, Sunghoon, Hyowon Ha, Hae-Gon Jeon, Stephen Lin, and In So Kweon. 2019. "Deep Depth from Uncalibrated Small Motion Clip." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, October. https://doi.org/10.1109/TPAMI.2019.2946806.
10. Kingma, Diederik P., and Max Welling. 2019. *An Introduction to Variational Autoencoders*.
11. Kiss-Illés, Dániel, Cristina Barrado, and Esther Salamí. 2019. "GPS-SLAM: An Augmentation of the ORB-SLAM Algorithm." *Sensors* 19 (22). https://doi.org/10.3390/s19224973.
12. Liu, Wei, Tao Wang, and Yachong Zhang. 2014. "A Relative Map Approach for Efficient EKF-SLAM." *Proceedings of 2014 IEEE Chinese Guidance, Navigation and Control Conference*. https://doi.org/10.1109/cgncc.2014.7007586.
13. Liu, Wenlei, Sentang Wu, Zhongbo Wu, and Xiaolong Wu. 2019. "Incremental Pose Map Optimization for Monocular Vision SLAM Based on Similarity Transformation." *Sensors* 19 (22). https://doi.org/10.3390/s19224945.
14. Mercy Rajaselvi Beaulah, P., D. Manjula, and Vijayan Sugumaran. 2018. "Categorization of Images Using Autoencoder Hashing and Training of Intra Bin Classifiers for Image Classification and Annotation." *Journal of Medical Systems* 42 (7): 132.
15. Sheng, Jinbo, Shunichi Tano, and Songmin Jia. 2011. "Mobile Robot Localization and Map Building Based on Laser Ranging and PTAM." *2011 IEEE International Conference on Mechatronics and Automation*. https://doi.org/10.1109/icma.2011.5985799.
16. Shen, Yuming, Li Liu, and Ling Shao. 2017. "Unsupervised Deep Generative Hashing." *Procedings of the British Machine Vision Conference 2017*. https://doi.org/10.5244/c.31.103.

17. Sumikura, Shinya, Mikiya Shibuya, and Ken Sakurada. 2019. “OpenVSLAM.” *Proceedings of the 27th ACM International Conference on Multimedia*. https://doi.org/10.1145/3343031.3350539.
18. Thrun, Sebastian, Wolfram Burgard, and Dieter Fox. 2005. *Probabilistic Robotics*. MIT Press.
19. Vlaminck, Michiel, Hiep Luong, and Wilfried Philips. 2018. “Have I Seen This Place Before? A Fast and Robust Loop Detection and Correction Method for 3D Lidar SLAM.” *Sensors* 19 (1). https://doi.org/10.3390/s19010023.
20. Wen, Jingren, Chuang Qian, Jian Tang, Hui Liu, Wenfang Ye, and Xiaoyun Fan. 2018. “2D LiDAR SLAM Back-End Optimization with Control Network Constraint for Mobile Mapping.” *Sensors* 18 (11). https://doi.org/10.3390/s18113668.

Biographical Sketch

Shaif Chowdhury received his B.Tech in Information Technology from Institute of Engineering and Management, Salt Lake, Kolkata, West Bengal, India.

After a brief stint at a Startup(ExamPreparationOnline) he worked at CSIR : Central Drug Research Institute as a Data Science intern. Followed by that he worked at IIT Kharagpur SRIC (National Digital Library) as a Project Fellow.

He is currently a PHD student at the Rochester Institute of Technology. His research interests include Computer Vision, Image Processing, Robotics and pattern Recognition.