Data-Efficient and Hardware Decentralized Visual SLAM

Jincheng Yu[1] and Feng Gao[1]

***Abstract*— Decentralized simultaneous localization and mapping (DSLAM) is essential to a multi-robot system, especially in environments lacking absolute positioning equipments like GPS. Visual based SLAM is a widely adopted solution in industry for its low cost and high flexibility. There are two essential components need to be efficiently deplyed on each agent: 1) Visual Odometr(VO) and 2) Place Recognition. However, both of these components require intensive computation and storage on embedded system. The place recognition task is usually done with CNN based methods. We adopt CNN as the VO to provide 6-D pose between different frames, for both intra-robot or inter-robot. Thus we can use the CNN accelerator based on FPGA to execute these two components. In this work, we propose a hardware-software co-design DSLAM framwork and use embedded FPGA to accelerator these two components.**

We evaluate our framework on the hardware platform Xilinx ZU9 SoC and we can perform DSLAM in real time on each agent. We also evaluate our system on publicly available dataset.

I. Introduction

In recent years, with the development of the hardware and algorithms, the capabilities of a single agent have been greatly improved. To further expand the capabilities of intelligent robots, using several robots can accelerate many tasks, such as localization, exploration, and mapping. As simultaneous localization and mapping(SLAM) is an essential component in many tasks, it is important to do SLAM across different robots in many multi-agent applications. The camera is a widely used snesor in SLAM for its rich information and low cost. However, in many scenarios, communication is limited, so that there is no server or an agent can stably collect all of the visual data from each robot.

Therefore, to reduce communication requirements, the previous work [1] proposes a data-efficient decentralized SLAM(DSLAM) system. The DSLAM frame in [1] is illustrated in fig 1(a). It makes improvements in three typical components in the DSLAM system: 1) Using ORB-SLAM [**?**] in stereo configuration as the visual odometry (VO) algorithm which provides basic intra-robot pose estimation. 2) Using NetVLAD [2] algorithm to do place recognition which relates the current observation to previous scenes and other robots. 3) Using distributed Gauss-Seidel algorithm [**?**] as the optimization back-end which optimizes the intra-robot position and fuses the inter-robot locations and maps. Each agent executes the ORB-SLAM which contains three steps for each input frame: feature extraction, feature matching and RANSAC. The NetVLAD method can encode the camera frame to a short vector which can be transformed to the server or a central agent with low communication cost.

[1]Electronic Engineering Department, Tsinghua University, Beijing, China
yjc16@mails.tsinghua.edu.cn

However, both ORB-SLAM and NetVLAD require tremendous computation and storage resources, and thus, the deployment of DSLAM on embedded system is challaged by the limited resources and power supply.

Though NetVLAD consumes huge computation, with the development of FPGA accelerators, we use the embedded CNN accelerator on FPGA [**?**] to perform NetVLAD for each frame. We also notice that there are also some previous works regression the 6-D pose directly from the input setero camera [**?**] or monocular camera [3], [4], With the development of CNN. We adopt Depth-VO-Feat [4] in DSLAM system to estimate the pose from the input monocular camera. Because Depth-VO-Feat is trained with stereo input frames and inferenced with monocular camera, the CNN method can provide absolute scale from monocular camera, and also be accelerated with our CNN accelerator. Thus we do not need to execute ORB-SLAM on embedded CPUs.

The proposed DSLAM framework is illustrated in fig 1(b). To make the DSLAM system more energy efficient and hardware friendly, we propose a novel hardware-software co-design DSLAM framework with the following contributions:

- We implement NetVLAD on an embedded SoC platform with CPU cores and FPGA fabric.
- We use an end-to-end CNN based method to estimate the 6-DoF pose between intra-robot successive frames and matched scenes between different robots.
- We demonstrate that our proposed hardware-software co-design decentralized SLAM system can achieve a similar accuracy with the current state-of-the-art DSLAM system without increase of communication.

The rest part of this article is orgnized as follows. Section II will give the basic idea of CNN based methods and the hardware architecture of embedded FPGA. Section III will detail the implementation of our hardware-software co-design DSLAM system. The experiment result will be given in Section IV. Section V will conclude this paper.

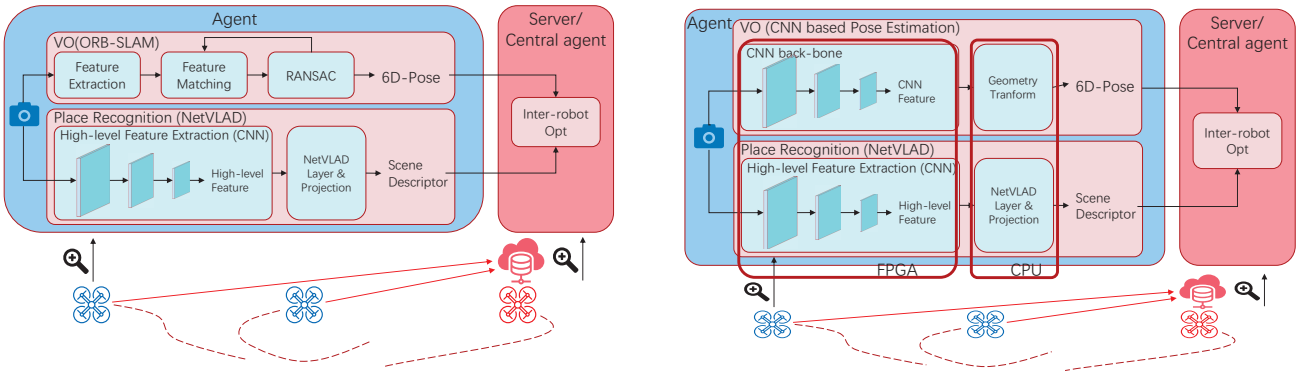
II. Background

A. CNN based methods in DSLAM

As described before, there are two essential components on each agent: 1) Visual Odometr(VO) and 2) Place Recognition.

B. Hardware architecture of Zync SoC

The Xilinx Zync Soc is a chip with ARM cores and FPGA fabric. The system is illustrated in section II-B. The ARM cores with an embedded Linux operation system are called Processing System (PS). The FPGA fabric is called Programmable Logic (PL). The peripherals like camera and



(a) DSLAM in [1]

(b) Our hardware-software co-design DSLAM.

Fig. 1. Overview of the DSLAM in [1] and our hardware-software co-design DSLAM. Each agent (blue drones) will send the result of 6-D pose estimation and scene descriptor to a server or a central agent (red server or drone in figure) to do inter-robot place recognition and optimization. We use CNN instead of feature points to do pose estimation so that we can use CNN accelerator to speed up the whole process.

communication unit (WiFi or others) are accessable with PS. The high-bandwidth on-chip AXI interface is used to communicate between PS and PL. PS and PL can also share the DDR to transfer large volume of data such as each frame of camera. Deephi CNN accelerator [**?**] is one of the state-of-the-art accelerators and is famous for high energy efficiency on various of CNN structure. We deploy the accelerator on the PL side of Zynq SoC.

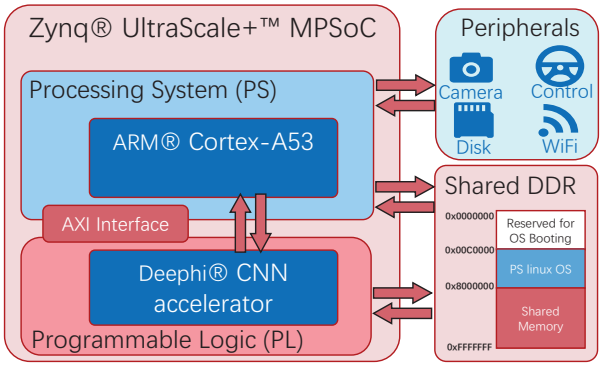


Fig. 2. Hardware architecture of Zynq SoC

III. Hardware-Software Co-design DSLAM

Our hardware-software co-design DSLAM system contains two essential improvements.

IV. Experiments

The experiment results shows our proposed DSLAM system can perform in real time and achieves similar accuracy with previous work.

V. Conclusion

Acknowledgment

This work is not supported by any fund.

References

[1] T. Cieslewski, S. Choudhary, and D. Scaramuzza, "Data-Efficient Decentralized Visual SLAM," *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2466–2473, 2018.

[2] R. Arandjelovic, P. Gronat, A. Torii, T. Pajdla, and J. Sivic, "NetVLAD: CNN Architecture for Weakly Supervised Place Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, pp. 1437–1451, 2017.

[3] R. Li, S. Wang, Z. Long, and D. Gu, "UnDeepVO: Monocular Visual Odometry Through Unsupervised Deep Learning," *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 7286–7291, 2018.

[4] H. Zhan, R. Garg, C. S. Weerasekera, K. Li, H. Agarwal, and I. Reid, "Unsupervised learning of monocular depth estimation and visual odometry with deep feature reconstruction," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.