

Design-Space Exploration of Cost-Efficient Point Cloud Processing on Embedded FPGAs

Keisuke Sugiura

August 2024

報告番号	㊦ 乙 第	号	氏 名	杉浦 圭祐
主 論 文 題 名 : Design-Space Exploration of Cost-Efficient Point Cloud Processing on Embedded FPGAs (組込み FPGA におけるコスト効率の良い点群処理の設計空間探索)				
(内容の要旨) 低コストな点群取得センサ (LiDAR など) の登場により、点群は様々なビジョン・ロボットタスクで使われる、標準的なデータ形式となった。点群ベースタスクは通常、移動ロボット上で動作するが、これらのロボットに搭載されるようなエッジデバイスの計算資源は、最新の手法をリアルタイム動作させるには不十分である。本論文は、エッジデバイス側の限られた処理能力と、点群ベースタスク (LiDAR SLAM、経路計画、レジストレーション) の高い計算要求とのギャップを埋めることを目的とする。 LiDAR SLAM に関しては、GEM (Greedy Endpoint Matching) と CSM (Correlative Scan Matching) とよばれる 2 つのスキャンマッチング手法を対象に、リソース効率のよい専用アクセラレータ (GEMCore、CSMCore) を提案する。これらは、小型 FPGA ボードの一つである Pynq-Z2 上に実装され、様々な SLAM システム (パーティクルフィルタ、グラフベース、スキャンマッチングベース SLAM) に統合される。IP コアによって、最新手法とも比肩する精度と、リアルタイム性能の双方を実現する。CSMCore は、スキャンマッチングおよびループ検出処理を最大 14.84 倍、18.92 倍高速化し、全体の性能を最大 4.67 倍改善している。 2D/3D 経路計画については、軽量の深層学習ベースの手法 P3Net (PointNet-based Path Planning Networks) と、組み込み FPGA (Xilinx ZCU104) を対象とした専用 IP コア P3NetCore を提案する。P3Net は、最近提案された MPNet アルゴリズムの改良版である。2D データセットにおいて、P3NetCore は、Nvidia Jetson と比べて最大 143.62 倍高速に動作する。消費電力は僅か 0.255W であり、GPU 搭載のワークステーションに対して最大 1278.14 倍の電力効率を実現している。P3Net は、MPNet に比べてニューラルネットのパラメータ数を 32.32 倍削減しつつ、経路計画の成功率を最大 28.2%改善し、最適に近い経路を出力できる。これらの性能改善は、IP コアの性能およびリソース消費の正確なモデル化と、徹底した設計空間探索によって達成されている。 点群レジストレーションに対しては、高速・高精度・ロバストなレジストレーションパイプラインを、組み込み FPGA (Xilinx ZCU104) 上に実現する。PointNet 特徴量に基づく Correspondence-free な (対応関係に依存しない) 手法、特に PointNetLK と ReAgent の 2 つに着目し、専用アクセラレータ (PointLKCore、ReAgentCore) を開発する。これらのアクセラレータは、Nvidia GeForce GPU と比較して最大 213.58 倍のエネルギー効率を達成している。古典手法よりも、点群間の大きなずれやノイズに対して頑健であるうえ、15ms 未満でほぼ最適な解に収束し、リアルタイム性能を実証している。				

Thesis Abstract

No. _____

Registration Number	☒ “KOU” ☐ “OTSU” No. _____ *Office use only	Name	Keisuke Sugiura
Thesis Title Design-Space Exploration of Cost-Efficient Point Cloud Processing on Embedded FPGAs			
Thesis Summary Point clouds have become a popular data format in a wide range of vision and robotic tasks, thanks to the development of low-cost sensors (e.g., LiDAR). While point cloud-based tasks are usually deployed on mobile robots, they often lack the computing power to run recent methods in real-time. This thesis aims to bridge the gap between the limited computing power of embedded devices and high computational complexity of three robotic applications: LiDAR SLAM, path planning, and point cloud registration. For 2D LiDAR SLAM, this thesis proposes two resource-efficient accelerators, GEMCore and CSMCore, to address the time-consuming scan-to-map matching based on GEM (Greedy Endpoint Matching) and CSM (Correlative Scan Matching). They are implemented on the Pynq-Z2 board and integrated into various SLAM methods, including particle filter-, graph-, and scan matching-based ones. The proposed design achieves real-time performance and high accuracy that is even comparable to the state-of-the-art. CSMCore effectively speeds up scan matching and loop closure detection by up to 14.84x and 18.92x, yielding up to 4.67x overall performance improvement. For 2D/3D path planning, this thesis proposes P3Net (PointNet-based Path Planning Networks), a lightweight deep learning-based method inspired by the recently-proposed MPNet, along with its dedicated IP core (P3NetCore) targeting embedded FPGA SoCs (Xilinx ZCU104). In the 2D case, P3NetCore runs up to 143.62x faster than Nvidia Jetson, consumes only 0.255W, and is 1278.14x more power-efficient than a GPU workstation. P3Net improves the success rate by up to 28.2% with 32.32x fewer parameters than MPNet, and plans near-optimal paths. These performance improvements are attained through accurate performance and resource modeling, as well as rigorous design-space exploration of the IP core. For 3D point cloud registration, this thesis develops a fast, accurate, and robust registration pipeline on embedded FPGAs. Two custom accelerator cores, PointLKCore and ReAgentCore, are designed for correspondence-free methods based on PointNet feature extraction (PointNetLK and ReAgent). On the Xilinx ZCU104, they achieve up to 213.58x better energy efficiency than an Nvidia GeForce GPU. The proposed cores are more robust to noise and large initial misalignments than classical methods and quickly find reasonable solutions in less than 15ms, demonstrating real-time performance.			