

무인이동체의 기능 동작 속도를 향상시키는 가변 구조를 갖는 가속기

김민지¹, 송원재², *이성주¹

¹정보통신학과 및 지능형 융합드론학과, 세종대학교

²전자정보통신공학과, 세종대학교

minji@itsoc.sejong.ac.kr, wonjae@itsoc.sejong.ac.kr, *seongjoo@sejong.ac.kr

Accelerator with variable structure that improves the functional operating speed of unmanned vehicles

Minji Kim¹, Wonjae Song², Seongjoo Lee^{1*}

¹Dept. of Information and Comm. Eng. and Coverage Engineering for Intelligent Drone,
Sejong Univ.

²Dept. of Electrical Eng., Sejong Univ.

요 약

현재 자율주행 이동체인 차량 및 드론이 발전함에 따라서 이와 관련된 연구도 활발하게 진행되고 있으며, 이에 따른 가속기 연구 또한 활발하게 진행되고 있다. 현재 다양하게 연구되고 있는 분야는 CPU와 GPU 그리고 다양한 종류의 FPGA가 설치 되어진 무인 이동체이다. 본 논문은 CPU와 GPU의 성능보다 뛰어난 DPU인 Alveo를 연구하였다. 계산량이 많고 복잡한 algorithm의 기능은 이런 기능을 동작했을 때의 속도가 무엇보다 중요하다. 그래서 CPU와 GPU 그리고 Alveo의 속도를 비교하여무인 이동체의 성능을 개선시키려고 했으며 이에 어울리는 가속기인 Alveo를 제안하였다.

I. 서 론

4차 산업 혁명으로 자율주행 이동체에 대한 관심이 나날이 증가하고 있는 가운데 무인 이동체의 기능이 동작하는 속도를 향상시키는 연구가 진행되고 있다. 무인 이동체의 기능이 작동되는 속도를 향상시키고자 CPU와 GPU 그리고 다양한 종류의 FPGA를 설치할 때도 있다[1][2]. 또한 현재 무인 이동체에 자율주행 등 복잡한 연산을 하는 기능들을 위해서도 성능 좋은 CPU나 GPU 또는 다양한 종류의 FPGA가 무인 이동체에 설치되고 있다[3][4]. 또한 한 가속기에 정착하여 연산 실행률을 향상시키는 연구가 진행되고 있으며, 가속기의 단점을 해결하고자 algorithm을 개선하는 연구 또한 다양하게 진행되고 있다.

본 논문에서는 가속기들의 종류 중 하나인 CPU와 GPU의 성능 비교에 더불어 DPU의 성능을 함께 비교해 어떤 가속기가 무인 이동체에 설치될 때 기능 동작 속도를 가장 향상시킬 수 있는지 알아볼 것이다. 모든 종류의 GPGPU가 많지만 말 그대로 범용이기 때문에 성능에 한계가 있었다. Xilinx의 DPU인 Alveo는 실리콘에 고정된 단위 프로세서로 구성된 GPU와 달리 단위 프로세서 수준에서 하드웨어를 재구성하기 때문에 단위 프로세서당 전력 효율 측면에서 장점이 있는 가속기이다. 본 논문에서는 간단한 계산, vector addition 방법으로 Alveo의 계산 속도를 기존 CPU 및 GPU와 얼마나 다른지 비교한다. 구현은 Intel의 CPU와 NVIDIA의 GPU 그리고 Xilinx의 DPU인 Alveo를 이용해 구현을 진행할 것이다.

II. 본론

본 논문에서는 GPGPU의 종류인 CPU, GPU 그리고 DPU의 계산 시간을 측정하면서 어떤 GPGPU가 가장 계산 속도가 뛰어난지 비교했다.

CPU, GPU 그리고 DPU인 Alveo의 계산 시간을 측정하기 위해 똑같은 프로그램을 이용했는데, 간단한 algorithm 중에서 vector addition을 사용했다. 서로 다른 배열을 더할 때 같은 index에 존재하는 값을 더해 연산하는 방식이다. 식으로 표현하면 다음과 같다.

$$C[n] = A[n] + B[n], (n = 0, 1, 2, \dots, m) \dots (1)$$

먼저 GPU의 계산 속도를 측정하기 위해서 Linux에서 Cuda 프로그램을 통해 계산 시간을 알아내었다. Cuda를 이용하기 위해서 .cu 파일을 생성해 코딩을 하고 test라는 파일을 생성했다. test라는 파일을 GPU 프로파일러를 이용해 실행하면 다음 그림에서 볼 수 있듯이, vector addition에서 사용된 모든 커널 및 메모리 복사본의 요약은 빠르게 확인할 수 있다. 그림 1과 같이 GPU의 vector addition 계산 시간은 약 180us이다.

다음으로 CPU와 DPU인 Alveo의 계산 속도를 측정하기 위해서 Linux에서 Vitis 프로그램을 통해 계산 속도를 알아내었다. Vitis 프로그램을 통해서 .cpp 파일에다가 코딩을 하고 build를 통해서 .xo 파일을 만들고 .xo

*교신저자 : 이성주

파일을 이용해서 Hardware 컴파일을 하여서 .xclbin 파일을 만든다. 그리고 Linux에서 .cpp 파일을 이용해 .exe 파일을 만든다. Alveo에 .xclbin을 다운로드 하고 .exe 파일을 이용해 .xclbin 파일을 실행한다. 이때 Linux의 debug 모드를 이용한다. debug 모드에서 실행을 시키면 아래의 그림과 같이 CPU와 DPU인 Alveo의 계산 시간이 다음과 같이 나온다.

그림 2와 같이 CPU의 vector addition 계산 시간은 약 23000us이고, Alveo의 계산 시간은 50us이다.

본 논문에서는 4,194,304개의 값을 가지고 각 vector들을 덧셈하여 CPU, GPU 그리고 Alveo의 계산 시간 차이를 비교하였다. 그리고 이를 통해서 Alveo의 연산 속도가 CPU보다는 약 460배 정도 빠르고 GPU보다는 약 3배 정도 빠른 것을 확인할 수 있었다. 그리고 이는 연산량이 많을수록 속도 차이가 커지는 것을 알 수 있었다. 밑의 그림은 16,777,216개의 값을 가지고 각 vector들을 덧셈하였을 때의 CPU, GPU 그리고 Alveo의 계산 시간이다.

그림 3, 4와 같이 CPU는 약 56327us이고, GPU는 약 721us이다. 마지막으로 Alveo는 약 40us이다. Alveo의 연산 속도가 CPU보다 약 1400배 정도 빠르고 GPU보다 약 18배 정도 빠른 것을 확인할 수 있다. 계산량이 커질수록 CPU, GPU와 Alveo의 계산 속도 차이가 커지는 것을 알 수 있다.

III. 결론

본 논문에서는 CPU, GPU와 Alveo의 연산 속도 차이로 인한 Alveo의 성능이 뛰어난 것을 보여주었다. Alveo는 CPU를 뛰어넘은 가속기인 GPU도 우월한 성능을 보여주었고 이는 복잡한 알고리즘이나 계산량이 많을수록 더 잘 보여주는 것을 알 수 있었다. 따라서 Alveo는 무인 이동체 기능의 속도를 빠르게 향상시키는데 기여할 수 있을 것으로 기대되어 진다.

```
==955== Profiling result:
Type      Time      Calls      Avg      Min      Max      Name
GPU activities: 61.82% 14.34ms 1 14.34ms 14.34ms 14.34ms [CUDA memcpy DtoH]
              37.41% 8.679ms 2 4.339ms 4.338ms 4.340ms [CUDA memcpy HtoD]
              0.77% 178.97us 1 178.97us 178.97us 178.97us vecAdd(double*, double*, double*, int)
API calls: 84.82% 151.60ms 3 50.53ms 67.88ms 151.40ms cudaMalloc
              15.64% 24.37ms 3 8.123ms 4.443ms 15.44ms cudaMemset
              1.15% 2.063ms 3 687.82us 204.16us 936.66us cudaFree
              0.23% 402.53us 1 402.53us 402.53us 402.53us cuDeviceGetAttribute
              0.11% 191.86us 101 1.899ms 197ms 81.35us cuDeviceGetAttribute
              0.03% 46.387us 1 46.387us 46.387us 46.387us cuDeviceGetName
              0.02% 44.249us 1 44.249us 44.249us 44.249us cuDeviceGetPCIBusId
              0.01% 9.9240us 1 9.9240us 9.9240us 9.9240us cuDeviceGetCount
              0.00% 1.8508us 3 616ns 248ns 1.2720us cuDeviceGet
              0.00% 1.1899us 2 594ns 230ns 959ns cuDeviceGet
              0.00% 363ns 1 363ns 363ns 363ns cuDeviceGet
```

그림 1. 4,194,304개 값의 vector addition GPU 계산 시간

```
(gdb) run
Starting program: /home/itsoc/workspace/vitis_rtl_kernel/test_rtl/Hardware/app.exe krnl_vadd.xclbin
[Thread debugging using libthread_db enabled]
Using host libthread_db library "/lib/x86_64-linux-gnu/libthread_db.so.1".
INFO: Found Xilinx Platform
INFO: Loading 'krnl_vadd.xclbin'
[New Thread 0x7fffe4987700 (LWP 884)]
[New Thread 0x7fffe1867000 (LWP 885)]
[New Thread 0x7fffe0985700 (LWP 886)]
[New Thread 0x7fffcffff700 (LWP 887)]
[New Thread 0x7fffc77fe700 (LWP 888)]
[New Thread 0x7fffc67fe700 (LWP 889)]
Alveo : 54359ns
CPU : 23079683ns
테스트가 ..... PASSED
[Thread 0x7fffc67fe700 (LWP 889) exited]
[Thread 0x7fffc77fe700 (LWP 888) exited]
[Thread 0x7fffcffff700 (LWP 887) exited]
[Thread 0x7fffe0985700 (LWP 886) exited]
[Thread 0x7fffe1867000 (LWP 885) exited]
[Thread 0x7fffe4987700 (LWP 884) exited]
```

그림 2. 4,194,304개 값의 vector addition CPU와 Alveo 계산 시간

```
==31812== Profiling result:
Type      Time      Calls      Avg      Min      Max      Name
GPU activities: 60.93% 59.700ms 1 59.700ms 59.700ms 59.700ms [CUDA memcpy DtoH]
              38.33% 37.555ms 2 18.777ms 18.674ms 18.881ms [CUDA memcpy HtoD]
              0.74% 721.16us 1 721.16us 721.16us 721.16us vecAdd(double*, double*, double*, int)
API calls: 61.09% 160.79ms 3 53.596ms 126.83us 160.53ms cudaMalloc
              37.72% 99.264ms 3 33.088ms 18.781ms 61.452ms cudaMemset
              0.93% 2.4452ms 3 815.08us 251.34us 1.1422ms cudaFree
              0.15% 403.34us 1 403.34us 403.34us 403.34us cuDeviceGetAttribute
              0.07% 195.54us 101 1.9360us 204ns 85.367us cuDeviceGetAttribute
              0.02% 44.701us 1 44.701us 44.701us 44.701us cuDeviceGetName
              0.01% 37.462us 1 37.462us 37.462us 37.462us cuDeviceGetPCIBusId
              0.00% 6.8460us 3 6846us 6.8460us 6.8460us cuDeviceGetCount
              0.00% 1.8909us 3 603ns 214ns 1.2860us cuDeviceGet
              0.00% 1.2390us 2 619ns 214ns 1.0250us cuDeviceGet
              0.00% 336ns 1 336ns 336ns 336ns cuDeviceGet
```

그림 3. 16,777,216개 값의 vector addition GPU 계산 시간

```
(gdb) run
Starting program: /home/itsoc/workspace/vitis_rtl_kernel/test_rtl/Hardware/app.exe krnl_vadd.xclbin
[Thread debugging using libthread_db enabled]
Using host libthread_db library "/lib/x86_64-linux-gnu/libthread_db.so.1".
INFO: Found Xilinx Platform
INFO: Loading 'krnl_vadd.xclbin'
[New Thread 0x7fffe4987700 (LWP 30261)]
[New Thread 0x7fffc77ff700 (LWP 30262)]
[New Thread 0x7fffc77fe700 (LWP 30263)]
[New Thread 0x7fffc67fe700 (LWP 30264)]
[New Thread 0x7fffc67fc700 (LWP 30265)]
[New Thread 0x7fffc5ffb700 (LWP 30266)]
Alveo : 40840ns
CPU : 56327994ns
테스트가 ..... PASSED
[Thread 0x7fffc5ffb700 (LWP 30266) exited]
[Thread 0x7fffc67fe700 (LWP 30265) exited]
[Thread 0x7fffc67fc700 (LWP 30264) exited]
[Thread 0x7fffc77fe700 (LWP 30263) exited]
[Thread 0x7fffc77ff700 (LWP 30262) exited]
[Thread 0x7fffc5ffb700 (LWP 30261) exited]
Inferior 1 (process 30258) exited normally
```

그림 4. 16,777,216개 값의 vector addition CPU와 Alveo 계산 시간

표 1. CPU, GPU, Alveo의 계산시간과 속도 비교

	CPU	GPU	Alveo
계산시간	56327us	721us	40us
속도	1	80	1400

ACKNOWLEDGMENT

본 연구는 정부(과학기술정보통신부)의 재원으로 한국연구재단(No.2020R1A2C1C007546) 및 정부(교육부)의 재원으로 한국연구재단의 이공분야 대학중점연구소지원사업(No. 2020R1A6A1A03038540)의 지원을 받아 수행되었으며, 검증은 위한 EDA관련 툴은 IDEC의 지원을 받았음.

참고 문헌

- [1] A. Fotouhi, M. Ding and M. Hassan, "Understanding autonomous drone maneuverability for Internet of Things applications," 2017 IEEE 18th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2017, pp. 1-6, doi: 10.1109/WoWMoM.2017.7974336.
- [2] R. Hossain, S. Magierowski and G. G. Messier, "GPU Enhanced Path Finding for an Unmanned Aerial Vehicle," 2014 IEEE International Parallel & Distributed Processing Symposium Workshops, 2014, pp. 1285-1293, doi: 10.1109/IPDPSW.2014.144.
- [3] P. A. Rad, D. Hofmann, S. A. Pertuz Mendez and D. Goehring, "Optimized Deep Learning Object Recognition for Drones using Embedded GPU," 2021 26th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), 2021, pp. 1-7, doi: 10.1109/ETFA45728.2021.9613590.
- [4] B. B. Kövari and E. Ebeid, "MPDrone: FPGA-based Platform for Intelligent Real-time Autonomous Drone Operations," 2021 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR), 2021, pp. 71-76, doi: 10.1109/SSRR53300.2021.9597857.