

저궤도 위성 온보드 시스템을 위한 CPU-GPU DVFS 기반 이미지 추론 및 코드 오프로딩 기법

김은수, 최평준, 김정환, kwakjeongho

대구경북과학기술원, 대구경북과학기술원, 대구경북과학기술원, 고려대학교

eskim1219@dgist.ac.kr, pyeongjun.choi@dgist.ac.kr, ghks9876@dgist.ac.kr,
jeonghokwak@korea.ac.kr

CPU-GPU DVFS-Based Image Inference and Code Offloading for LEO Satellite Onboard Systems

Kim Eun Su, Choi Pyeong Jun, Kim Jeong Hwan, Kwak Jeong Ho
DGIST., DGIST., DGIST., Korea Univ.

요약

저궤도 위성은 넓은 커버리지와 높은 기동성을 바탕으로 실시간 지구 관측 및 고해상도 이미지 수집에 활용되고 있다. 그러나 이러한 시스템에서 높은 정확도의 이미지 추론과 빠른 전송을 동시에 보장하기에는 제한된 온보드 연산 자원과 빠르게 변화하는 위성-지상 간 채널 상태와 같은 여러 제약으로 인해 상당한 어려움이 존재한다. 기존 연구들은 이러한 문제를 해결하기 위해 동적 이미지 인코딩과 동적 코드 오프로딩을 각각 독립적으로 최적화해 왔으나, 두 요소 간의 CPU/GPU 연산 자원의 동적 스케일링 및 스케줄링을 함께 고려하지 않아 전체적인 성능 향상에 한계가 있었다. 본 논문에서는 이러한 한계를 극복하기 위한 AERO 알고리즘을 제안한다. AERO는 동적 채널 환경에 적응적으로 이미지 인코딩, 코드 오프로딩 정책, 그리고 CPU/GPU 클럭 주파수를 동적으로 조정하고, 두 이중 프로세서를 교차적으로 스케줄링 하고 평균 추론 정확도 기준을 만족시키면서 위성 단말의 평균 전력 소비를 최소화하는 것을 목표로 한다. 또한, 실제 위성 OBP (OnBoard Processor)의 성능과 유사한 모바일 단말에서 측정한 인코딩 레이트-추론 정확도 관계 및 인코딩 레이트에 따른 인코딩 시간에 대한 사전 실험 데이터를 시뮬레이션에 반영하여 현실성을 높였다. 시뮬레이션 결과, 제안하는 AERO 알고리즘이 기존 알고리즘 대비 동일한 인코딩 통합 지연 시간 및 추론 정확도 대비 평균 전력을 최대 18% 절감할 수 있음을 보였다.

I. 서론

고해상도 지구 관측 영상은 재난 모니터링, 해양 감시, 농업 분석 등 다양한 분야에서 핵심 데이터로 활용되고 있으며, 특히 기후 변화, 국방 감시, 스마트 농업 등 광범위한 커버리지를 요구하는 응용에서 수요가 급증하고 있다. 이에 따라 저궤도 위성은 방대한 시각 데이터를 전 지구적으로 수집하는 차세대 위성 서비스의 핵심 플랫폼으로 주목받고 있다. 그러나 SAR 위성이나 고해상도 광학 센서와 같은 첨단 장비는 막대한 연산 자원과 통신 대역폭을 필요로 하며, 이는 제한된 위성 전력을 빠르게 소모시킨다. 특히 위성은 지상 시스템과 달리 배터리 충전이 제한적이어서 효율적인 전력 관리가 서비스 지속 가능성과 품질을 위해 필수적이다.

기존 연구에서 대표적인 접근법은 로컬 추론, 인코딩, 오프로딩이다. 로컬 추론은 채널 상태가 좋지 않아 전송전력이 급격히 증가할 때 유용하지만, 복잡한 연산을 수행할 경우 GPU 사용량이 늘어나 연산전력이 크게 소모된다. 인코딩은 전송전력을 절감할 수 있지만, 높은 인코딩율은 전력소모를 증가시키고 이미지 품질 저하로 인해 추론 정확도가 떨어진다. 오프로딩은 지상 서버의 풍부한 연산 자원을 활용할 수 있으나, 채널 변동성과

지연, 통신 전력 소모가 크다. 따라서 단일 접근만으로는 연산속도, 정확도와 전송전력을 동시에 최적화하기 어렵다. 이를 해결하기 위해 본 연구에서는 AERO (Adaptive Encoding and Inference Optimization) 알고리즘을 제안한다. AERO는 Lyapunov 최적화 기법을 적용하여 매 타임 슬롯마다 오프로딩 정책, CPU/GPU 클럭 주파수, 이미지 인코딩율을 통합적으로 제어하여 추론 정확도와 지연 제약 조건을 만족시키면서 장기 평균 전력소비를 최소화한다. 특히 AERO는 동적 채널 상태와 처리 큐 정보를 반영하여 인코딩율을 적응적으로 조정하고, CPU와 GPU를 상호 교차적으로 스케줄링하는 최초의 위성 온보드 통합 제어 프레임워크라는 점에서 차별성을 가진다.

이 논문의 기여를 정리하면 다음과 같다.

- 1) 우리는 평균 추론 정확도 제약 및 평균 전송지연을 유지하면서 평균전력 소비를 최소화하는 새로운 최적화 문제를 수립하고, 이 문제를 해결하기 위한 최적 통합 제어 알고리즘 AERO를 제안하였다.
- 2) 본 연구는 위성 온보드 시스템에서 CPU와 GPU를 상호 교차적으로 스케줄링하고, 이미지 추론과 인코딩을 통합적으로 제어하는 최초의 연구이다.
- 3) 우리는 실제 디바이스 측정값을 반영한 시뮬레이션을

통해 기존 알고리즘 대비 최대 18%까지 전력 성능을 향상시키는 것을 확인하였다.

II. 본론

실험 세팅. 본 연구의 사전실험은 삼성 갤럭시 S21 울트라 단말에서 수행되었으며, COCO dataset의 이미지를 사용하였다. YOLOv8m 모델을 위성 온보드 환경에, YOLOv8l 모델을 지상 서버 환경에 각각 적용하여 압축률 변화가 추론 정확도에 미치는 영향을 분석하였다. 또한, JPEG 압축 과정에서 CPU 클럭 주파수를 변화시키고, Quality Factor(QF) 값을 조정하여 인코딩 시간의 변화를 측정하였다.

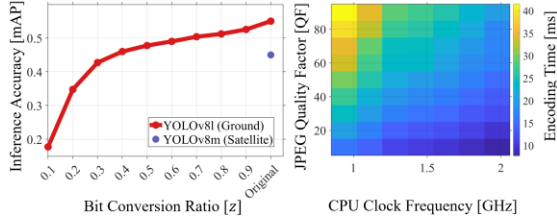


Fig. 1. 추론 정확도 및 인코딩 시간 그래프

실험 결과. Fig. 1의 왼쪽 그래프는 인코딩을 z 에 따른 추론 정확도 변화를 나타낸다. z 이 높을수록 이미지 품질 저하로 인해 정확도가 감소하였다. 또한, Fig. 1의 오른쪽 그래프는 CPU 클럭 주파수와 QF에 따른 JPEG 인코딩 시간을 보여준다. CPU 클럭 주파수가 높을수록 인코딩 시간이 단축되었으며, QF가 낮아질수록(즉, 인코딩률이 높을수록) 이미지 크기 감소로 인해 인코딩 속도가 빨라졌다.

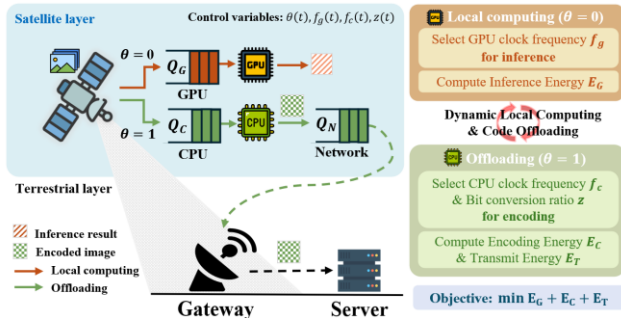


Fig. 2. 제안하는 AERO 프레임워크

시스템 모델. 우리는 Fig. 2와 같이 단일 위성과 다수의 지상 게이트웨이가 존재하는 시스템을 고려한다. 저궤도 위성은 고정된 궤도를 따라 지구를 공전하며, 가시 범위 내의 게이트웨이와 통신 링크를 형성하여 데이터를 송수신한다. 지상 계층에는 게이트웨이가 균일하게 배치되어 핸드오버를 통해 끊김 없는 연결을 유지한다. 이 프레임워크는 슬롯 기반 시스템으로 동작하며, 매 슬롯마다 위성은 두 가지 동작 중 하나를 선택한다. 1. 로컬 추론: 온보드 GPU를 사용해 수집한 이미지를 직접 추론하고, 결과만 지상으로 전송한다. 2. 오프로딩: 온보드 CPU를 사용해 이미지를 JPEG 인코딩한 뒤, 지상 서버로 전송하여 지상 서버 측에서 추론을 수행한다. 우리의 목표는 오프로딩 정책 θ 와 CPU/GPU 클럭 주파수 f_g, f_c 그리고 압축률 z 를 동적으로 제어하여 LEO 위성 온보드의 평균 총 사용전력(추론 + 인코딩 + 전송)을 최소화하는 것이다.

알고리즘. 우리는 평균 총 에너지 최소화 문제를 실시간으로 해결하고 큐 안정성을 보장하기 위해 Lyapunov 기반 제어 알고리즘인 AERO를 개발하였다. 제안하는 AERO 알고리즘은 Lyapunov 최적화 프레임워크를 기반으로 하여 매 시각 슬롯마다 오프로딩 여부, GPU/CPU 주파수, 압축률을 동적으로 선택한다. 먼저 GPU 주파수는 에너지 소비와 큐 지연을 동시에 고려해 최적 값을 결정하고, CPU 주파수와 압축률은 에너지, 큐 길이, 다운로드 채널 상태를 반영한 비용 함수를 최소화하여 선택한다. 이후 로컬 처리와 오프로딩 비용을 각각 계산한 뒤, 두 값을 비교해 더 낮은 비용을 갖는 방식으로 결정한다. 이를 통해 AERO 알고리즘은 에너지-지연-정확도 간의 트레이드오프를 효과적으로 조율한다.

성능 평가. Fig. 3은 제안하는 AERO 알고리즘의 평균 에너지 소비와 평균 큐 지연간의 관계를 보여준다. 특히, AERO는 오프로딩할 때 압축을 하지 않는 No encoding 기법에 비해 동일한 큐 백로그에서 약 18% 에너지 절감을 절감하였다. 이를 통해 AERO는 위성 환경에서 에너지-지연 간 균형을 효과적으로 달성하는 알고리즘임을 알 수 있다.

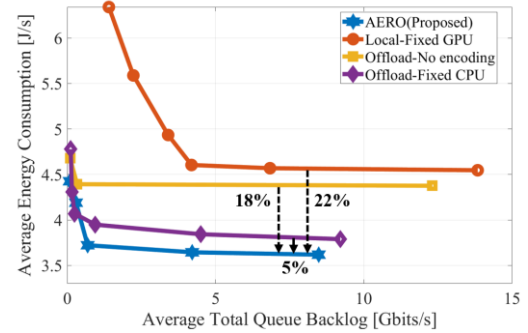


Fig. 3. 알고리즘별 Trade-off 그래프

III. 결론

본 논문에서는 저궤도 위성 환경에서 제한된 연산 및 전력 자원을 효율적으로 활용하기 위해, CPU/GPU DVFS 제어, 이미지 인코딩을 제어, 오프로딩 결정을 동적으로 최적화하는 AERO 알고리즘을 제안하였다. 본 연구는 차세대 저궤도 위성 기반 협력 네트워크에서 온보드 자원 관리 및 추론 성능 최적화에 기여할 수 있을 것으로 기대된다.

ACKNOWLEDGMENT

이 논문은 2025년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.RS-2022-II220704, 초고속 이동체 지원을 위한 3D-NET 핵심 기술 개발)

참고 문헌

[1] Du, K. et al. "Earth+: On-Board Satellite Imagery Compression Leveraging Historical Earth Observations," Proceedings of the 30th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Vol.1, pp. 361-376 (2025).