

저전력 영상인식 시스템을 위한 온디바이스 심층신경망 기반 저전력 통신 기법

임재봉*, 백윤주°

Low-Power Communication Method using On-Device Deep Neural Network for Low-Power Image Recognition System

Jaebong Lim*, Yunju Baek°

요약

최근 사물인터넷 기술의 발달과 함께 저전력 영상인식 시스템이 활발하게 연구되고 있다. 기존에는 저전력 영상인식 시스템에서 높은 정확도의 영상인식 심층신경망을 활용하기 위하여, 장치에서는 이미지를 촬영만 하고 서버에서 이미지를 전달받아 영상인식 심층신경망을 활용하였다. 그러나 이미지 전달을 위한 장치와 서버 간 지속적인 통신으로 전력소모가 큰 문제가 있다. 따라서 본 논문에서는 온디바이스 심층신경망 기반 저전력 통신 기법을 제안하여 기존 기법의 통신 전력소모 문제를 개선하였다. 제약적인 자원의 장치에서 온디바이스 심층신경망을 활용하여, 영상인식 결과가 관심이미지인 경우에만 이미지를 전송함으로써 전력소모를 최적화하였다. 또한 제안 기법에 대한 전력소모 모델을 제시하고 기존 기법 대비 제안 기법의 전력소모와 동작시간을 평가하였다. Caltech 비디오 데이터셋트를 활용하여 실험한 결과, 제안 기법은 기존 기법 대비 전력소모가 54.4% 감소함을 보였다.

Key Words : Image Recognition System, Low-Power Communication, On-Device AI, Deep Neural Network, Resource-Constrained Device, Power Consumption Model

ABSTRACT

In recent years, low-power image recognition system has been actively studied with the evolution of IoT technology. In order to manipulate a deep neural network, which has high accuracy for image recognition, the device only captures images and transmits the image to the server and then performs image recognition on the server. However, there is a problem in that power consumption is large due to communication for continuous image transmission. In this paper, we propose low-power communication method using on-device deep neural network to solve the communication power consumption problem of image recognition systems. Using on-device deep neural network from the constrained resource device, we optimized the power consumption by transmitting the image only when the image recognition result is the image of interest. Also, we propose a power consumption model for the proposed method. And, we evaluate the power consumption and lifetime of the proposed method compared to baseline method. Experimental results show that the proposed method reduces power consumption by 54.4% compared to the baseline method.

* 이 연구는 2019년도 산업통상자원부 및 산업기술평가관리원(KEIT) 연구비 지원에 의한 연구임(20000515)

• First Author : Pusan National University Dept. of Electrical and Computer Engineering, jaebong.lim@eslab.re.kr, 학생회원

° Corresponding Author : Pusan National University Dept. of Electrical and Computer Engineering, yunju@pusan.ac.kr, 종신회원
논문번호 : 201905-059-C-RE, Received April 30, 2019; Revised June 19, 2019; Accepted July 2, 2019

I. 서 론

최근 심층신경망은 영상인식 분야에서 기존 기법에 비해 높은 정확도를 보여 다양한 장치와 응용에 널리 활용되고 있다^[1]. 심층신경망은 기존 기계학습 기법과 달리 데이터로부터 분류기뿐만 아니라 특징을 학습하여 높은 정확도를 보인다. 그러나 심층신경망을 이용하기 위해서는 특징 공간이 큰 특성으로 인해 많은 저장 공간과 높은 계산 성능을 요구하는 문제가 있다. 따라서 저장 공간과 계산 성능이 부족한 장치에서는 별도 서버를 활용하는 클라우드 컴퓨팅을 기반으로 심층신경망을 활용할 수 있다^[2]. 클라우드 컴퓨팅은 장치에서 수집한 영상을 서버에 전송하고 서버에서 심층신경망을 활용한 후 그 결과를 장치에 다시 전송하는 방식으로 동작한다. 그러나 클라우드 컴퓨팅은 서버와 장치 간의 통신으로 인해 전력소모가 큰 문제가 있다^[3]. 따라서 웨어러블 장치, 자동차 및 해양 장치 등 배터리를 전력원으로 하는 장치에서는 전력소모가 클 경우 동작 시간에 제약을 받는 문제가 있다.

클라우드 컴퓨팅 기반 영상인식 시스템의 전력소모 문제를 해결하기 위해서는 온디바이스 심층신경망^[4]을 활용한 저전력 영상인식 시스템의 연구가 필요하다. 전력소모의 원인인 서버와 장치 간 통신 없이 장치 내에서 동작하는 영상인식 심층신경망을 활용할 경우 전력소모를 개선할 수 있다. 영상인식 심층신경망은 많은 저장 공간과 높은 계산 성능을 가진 서버 또는 워크스테이션에서 활용하는 것이 일반적이거나 최근 제한적인 자원을 가진 장치에서 활용하기 위한 연구가 활발하다^[5]. 즉 서버와의 통신 없이 온디바이스에서 심층신경망을 활용할 수 있다. 그러므로 저전력 영상인식 시스템을 위하여 제한적인 자원을 가지는 장치에서 심층신경망을 활용하여 영상인식을 수행함으로써 통신으로 인한 전력소모를 줄일 수 있다.

온디바이스 심층신경망을 활용할 경우 다양한 응용에서 서버와의 통신을 최소화할 수 있다. 지능형 감시 카메라^[6]는 접근하는 사람을 인식했을 경우 관리자에게 실시간 알림을 전달한다. 이를 위하여 클라우드 컴퓨팅 기반의 영상인식 시스템을 활용할 경우 이미지를 서버로 지속해서 전송해야 하므로 장치의 전력소모가 크다. 그러나 온디바이스 심층신경망 기반의 영상인식 시스템을 활용할 경우 장치에서 사람을 인식했을 때만 알림과 함께 이미지를 전송하고 나머지의 경우 통신을 수행하지 않아 전력소모를 줄일 수 있다. 따라서 감시 카메라, 웨어러블 카메라, 홈 카메라, 차량 카메라, 해상 카메라 등에서는 온디바이스에서 영

상인식을 통하여 관심이미지만을 전송함으로써 전력소모를 줄일 수 있다^[7].

따라서 본 논문에서는 저전력 영상인식 시스템을 위한 온디바이스 심층신경망 기반 저전력 통신 기법을 제안하였다. 저전력 영상인식 시스템에서 온디바이스 심층신경망 기반 영상인식을 수행함으로써 관심이미지만을 서버에 전송하여 전력소모를 개선하였다. 그리고 기존 저전력 기법과 결합하여 전력소모를 최적화하고 이에 대한 전력소모 모델을 제안하였다. 그리고 구현한 MCU 장치에서 기존 기법과 제안 기법의 전력소모를 비교하여 평가하였다.

II. 관련 연구

이 장에서는 저전력 영상인식 시스템의 구성과 기존의 저전력 동작 기법에 대해 살펴본다. 먼저 기존 고성능 영상인식 시스템과 저전력 영상인식 시스템과의 차이를 설명하고, 다음으로 영상인식 시스템에서 전력소모 문제를 발생시키는 원인인 이미지센서 전력소모와 통신 전력소모에 대해 살펴본다^[8]. 기존의 저전력 동작 기법들은 저전력 영상인식 시스템에서 이미지센서에 의한 전력소모가 큰 문제를 해결하기 위하여 영상인식 시스템의 동작을 최적화하는데 중점을 두고 있다^[9-13]. 본 논문에서는 이와 더불어 최근 활발히 연구되고 있는 온디바이스 심층신경망을 활용함으로써 저전력 영상인식 시스템의 통신 전력소모 문제를 해결하고자 한다. 따라서 최근 연구되고 있는 온디바이스 심층신경망을 활용한 영상인식 연구에 대해 살펴본다.

2.1 저전력 영상인식 시스템

기존 고성능 영상인식 시스템은 영상인식 심층신경망을 가속할 수 있는 고성능의 장치를 활용하여 실시간으로 영상인식을 수행할 수 있다. 그러나 장치의 전력소모가 매우 크므로 배터리를 전력원으로 하는 시스템에서는 활용하는데 어려움이 크다. 최근 모바일 장치의 성능 향상과 심층신경망의 경량화로 심층신경망 기반 모바일 영상인식 시스템에 대한 연구가 활발하다^[14-16]. 그러나 전력소모 개선의 한계로 대용량 배터리의 활용에도 동작시간이 제약되는 문제가 있다. 따라서 이러한 문제를 해결하기 위하여 저전력의 Micro Controller Unit (MCU)를 활용한 저전력 영상인식 시스템에 대한 연구가 활발하다^[17]. 이를 활용할 경우 심층신경망 계산에 따른 전력소모를 개선할 수 있다. 이와 더불어 저전력 영상인식 시스템에서 이미

지센서로 인한 전력소모 문제를 해결하기 위한 연구가 활발하다. 대표적인 연구로 주기적 스케줄링 기법과 다중 MCU 기법이 있다.

주기적 스케줄링 기법^{9,10)}은 이미지센서에 의한 전력소모를 개선하기 위해서 응용에 적합한 특정 주기마다 이미지를 촬영하는 기법이다. 비디오를 촬영하거나 연속적인 이미지를 촬영할 경우 이미지센서는 전력소모가 큰 문제가 있다. 따라서 이 기법은 주기 T마다 Wake-up하여 이미지를 촬영하고 나머지 시간동안 저전력의 Sleep 상태로 대기한다. 응용에 적합한 주기 T를 선정함으로써 저전력 상태인 Sleep의 비율 조절하여 전력소모를 감소시킬 수 있다.

다중 MCU 기법¹¹⁻¹³⁾은 여러 개의 MCU를 활용하여 동작을 최적화함으로써 전력소모를 최소화하는 기법이다. 전력소모와 필요한 컴퓨팅자원에 따라 MCU별 동작을 분리하고 상황에 맞게 동작시킴으로써 전력소모를 최적화한다. 예를 들어 주 MCU의 단순 저전력 동작 수행과 보조 MCU의 센서 제어 및 복잡한 연산 수행으로 두 개 MCU의 동작을 분리하고, 저전력인 주 MCU가 보조 MCU의 전원을 제어함으로써 전력소모를 최적화 할 수 있다.

2.2 온디바이스 심층신경망

최근 효율적인 영상인식 심층신경망에 대한 연구가 활발하다. 기존 영상인식 심층신경망의 경우 높은 정확도를 위하여 모델의 크기가 매우 크고 연산 시간이 긴 문제가 있다. 이러한 문제를 해결하기 위하여 심층신경망 모델의 크기를 줄이고 연산시간을 줄이기 위한 연구가 활발하다. 대표적으로 MobileNet¹⁴⁾ 심층신경망은 Depthwise Separable Convolution을 도입하여 정확도 대비 모델 크기 및 연산 시간을 대폭 향상시켰다. 또한 SqueezeNet¹⁵⁾ 및 ShuffleNet¹⁶⁾ 등 다양한 효율적인 심층신경망들이 연구되고 있다.

MCU는 다양한 임베디드 응용에서 활용되는 저전력 프로세서로 컴퓨팅 자원의 제약이 크므로 심층신경망을 활용하는데 어려움이 있다. 그러나 최근 효율적인 심층신경망의 연구와 MCU의 성능 향상으로 MCU에서 심층신경망 활용하고자 하는 시도가 활발하다. 특히 MCU 중 가장 널리 사용되고 있는 구조인 ARM Cortex-M 구조의 경우 심층신경망 가속을 위한 다양한 명령어가 지원되고 있다. 특히 심층신경망 가속을 위하여 병렬연산을 지원하는 CMSIS-NN¹⁷⁾ 라이브러리를 활용할 경우 MCU에서 심층신경망 추론 시간을 대폭 감소시킬 수 있다.

III. 본 론

저전력 영상인식 시스템을 위해서는 이미지센서의 전력소모와 통신을 위한 전력소모를 최소화하여야 한다. 본 논문에서는 온디바이스 심층신경망 기반 저전력 통신 기법을 제안하고 기존 기법과 결합함으로써 이미지센서와 통신에 의한 전력소모를 개선하였다. 제안하는 저전력 통신 기법은 온디바이스 심층신경망을 활용하여 제약적인 자원의 장치에서 영상인식을 수행하고 관심이미지만을 서버에 전송함으로써 통신으로 인한 전력소모를 개선하였다.

본 장에서는 먼저 클라우드 컴퓨팅 기반 저전력 영상인식 시스템에 주기적 스케줄링 기법과 다중 MCU 기법을 적용한 기존 방법에 대해 설명한다. 이어서 기존 기법과 결합한 온디바이스 심층신경망 기반 저전력 통신 기법을 제안한다. 마지막으로 기존 기법과 제안 기법의 전력소모 모델을 제시한다.

3.1 온디바이스 심층신경망 기반 저전력 통신 기법

클라우드 컴퓨팅 기반 저전력 영상인식 시스템은 영상인식을 위하여 서버에 지속적으로 이미지를 전송할 필요가 있다. 따라서 시스템은 이미지를 촬영 후 게이트웨이에 이미지를 전송하고 게이트웨이는 다시 서버로 이미지를 전송한다. 그림 1 (a)와 같이 이미지 MCU는 초기화 후 이미지를 촬영하고 BLE 트랜시버를 활용하여 게이트웨이에 이미지를 전송한다. 클라우드 컴퓨팅 기반 저전력 영상인식 시스템은 영상인식을 위하여 이미지를 지속해서 전송해야 하므로 전력소모가 큰 문제가 있다.

이러한 문제점을 해결하기 위하여 클라우드 컴퓨팅 기반 영상인식 시스템에 주기적 스케줄링 기법을 적용하여 전력소모를 개선할 수 있다. 주기적 스케줄링 기법은 그림 1 (b)와 같이 주기 T마다 이미지를 촬영하고 이미지를 전송하는 기법으로 나머지 시간동안은 저전력의 Sleep 상태로 대기한다. 응용에 적합한 주기 T를 선정할 경우 성능은 유지하며 전력소모를 줄일 수 있다.

다중 MCU 기법은 BLE 트랜시버를 대신하여 BLE MCU를 적용하여 전력소모를 최적화할 수 있다. BLE MCU는 BLE 통신을 활용하여 이미지를 전송하고 이미지 MCU는 비교적 전력소모가 큰 이미지 촬영과 영상인식을 수행하도록 역할을 분리한다. 그리고 BLE MCU는 주기적으로 이미지 MCU의 전원을 제어하여 이미지를 촬영하고 영상인식을 수행하도록 한다. 이미지 MCU의 Sleep 동안의 전력소모 보다 BLE MCU

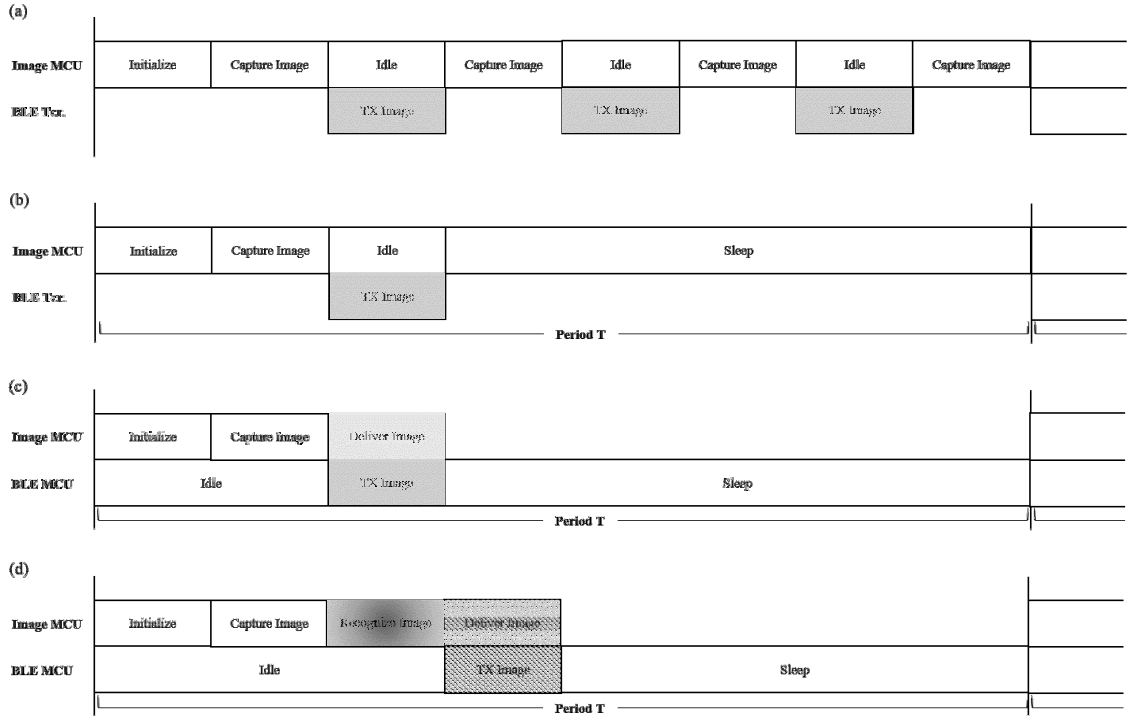


그림 1. (a) 클라우드 컴퓨팅 기반 저전력 영상인식 시스템의 동작 다이어그램 (b) 주기적 스케줄링 기법을 적용한 동작 다이어그램 (c) 주기적 스케줄링 기법과 다중 MCU 기법을 적용한 동작 다이어그램 (d) 제안하는 온디바이스 심층신경망 기반 저전력 통신 기법의 동작 다이어그램
 Fig. 1. (a) Sequence diagram of low-power image recognition system based on cloud computing (b) Sequence diagram with periodic scheduling method (c) Sequence diagram with periodic scheduling and multiple MCU method (d) Sequence diagram of proposed low-power communication method using on-device neural networks

의 Sleep 동작의 전력소모가 더 적으므로, 다중 MCU 를 도입함으로써 전력소모를 개선할 수 있다. 그림 1 (c)와 같이 BLE MCU는 주기 T마다 이미지 MCU의 전원을 인가하여 이미지를 촬영하고 이미지를 전달받아 게이트웨이로 이미지를 전송한다.

클라우드 컴퓨팅 기반 영상인식 시스템은 주기적 스케줄링 기법과 다중 MCU 기법을 활용하여 이미지 센서에 의한 전력소모 문제는 개선할 수 있지만 이미지 전송으로 인한 통신 전력소모가 여전히 큰 문제가 있다. 따라서 본 논문에서는 온디바이스 심층신경망 기반 저전력 통신 기법을 제안하고 이를 주기적 스케줄링 기법 및 다중 MCU 기법과 함께 적용하였다. 제안하는 온디바이스 심층신경망 기반 저전력 통신 기법은 그림 1 (d)와 같이 이미지 MCU에서 이미지를 촬영한 후 온디바이스 심층신경망을 활용하여 영상인식을 수행하고 관심이미지인 경우에만 이미지를 서버로 전송한다. 온디바이스 심층신경망을 활용한 영상인식은 BLE 통신을 통한 이미지 전송보다 전력소모가 낮다. 따라서 관심이미지의 비율이 낮을 경우 전력소

모를 대폭 개선할 수 있다.

제안 기법은 다양한 응용을 위한 저전력 영상인식 시스템의 전력소모를 개선한다. 지능형 감시 카메라를 위한 영상인식 시스템의 경우 접근하는 사람을 인식했을 경우 관리자에게 실시간 알림을 전달한다. 이러한 응용의 경우 전체 동작시간 동안 관심이미지의 비율이 낮은 특징 있다. 그러므로 제안 기법은 장치에서 사람을 인식했을 때만 알림과 함께 이미지를 전송하고 나머지의 경우 통신을 수행하지 않아 기존 기법에 비해 전력소모를 줄일 수 있다. 따라서 감시 카메라, 웨어러블 카메라, 홈 카메라, 차량 카메라, 해상 카메라 등에서는 제한 기법을 활용하여 관심이미지만을 서버에 전송함으로써 전력소모를 줄일 수 있다.

3.2 기존 기법 및 제안 기법의 전력소모 모델

그림 2 (a)는 주기적 촬영기법과 다중 MCU 기법을 적용한 기존 기법의 전력소모 모델이다. 이미지 MCU는 이미지센서를 초기화하고 이미지를 촬영한 후 촬영한 이미지를 BLE MCU에 전달한다. 그러므로

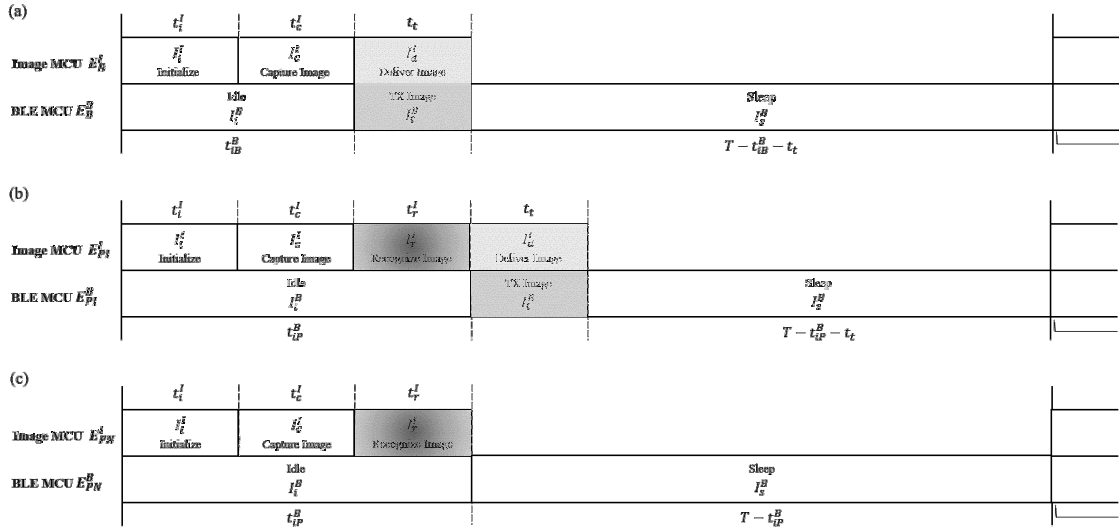


그림 2. (a) 기존 기법의 전력소모 모델 (b) 관심이미지일 경우 제안 기법의 전력소모 모델 (c) 관심이미지가 아닐 경우 제안 기법의 전력소모 모델
 Fig. 2. (a) Power consumption model of baseline method (b) Power consumption model of proposed method for images of interest (c) Power consumption model of proposed method for images of non-interest

이미지 MCU의 동작 별 전류 및 시간은 초기화 과정에서 I_i^I 및 t_i^I이고 촬영 과정에서 I_c^I 및 t_c^I이며 이미지를 전송하는 과정에서 I_d^I 및 t_t이다. 따라서 이미지 MCU의 에너지소모량 E_B^I은 이미지 MCU의 전압이 V일 때 식 1과 같다.

$$E_B^I = V^I(I_i^I t_i^I + I_c^I t_c^I + I_d^I t_t) \quad (1)$$

BLE MCU는 이미지 MCU가 동작할 동안 대기하다가 이미지를 전송받은 후 BLE 무선 통신을 이용하여 이미지를 전송한다. 그러므로 BLE MCU의 동작 별 전류 및 시간은 Sleep 중에 I_s^B 및 t_{IB}^B이고 이미지를 전송하는 과정에서 I_s^B 및 t_t이며 Sleep 중에서 I_s^B 및 T - t_{IB}^B - t_t이다. 따라서 BLE MCU의 에너지소모량 E_B^B은 BLE MCU의 전압이 V^B일 때 식 2와 같다. 그림 2 (a)의 이미지 MCU와 BLE MCU에 대한 평균 전력소모 P_B는 식 3과 같다.

$$E_B^B = V^B(I_s^B t_{IB}^B + I_t^B t_t + I_s^B (T - t_{IB}^B - t_t)) \quad (2)$$

$$P_B = (E_B^I + E_B^B) / T \quad (3)$$

제안 기법의 전력소모 모델의 경우 관심이미지가 촬영된 경우와 관심이미지가 아닌 경우로 나뉜다. 먼

저 그림 2 (b)와 같이 관심이미지가 촬영된 경우, 이미지 MCU는 이미지센서를 초기화하고 이미지를 촬영한 후 촬영한 이미지에 대하여 영상인식을 수행한다. 그리고 촬영한 이미지는 관심이미지이므로 BLE MCU에 전달한다. 그러므로 이미지 MCU의 동작 별 전류 및 시간은 초기화 과정에서 I_i^I 및 t_i^I이고 촬영 과정에서 I_c^I 및 t_c^I이고 영상인식 과정에서 I_r^I 및 t_r^I이며 이미지를 전송하는 과정에서 I_d^I 및 t_t이다. 따라서 이미지 MCU의 에너지소모량 E_{PI}^I은 이미지 MCU의 전압이 V^I일 때 식 4와 같다. 그리고 관심이미지의 경우 이미지 MCU의 에너지소모량은 기존 기법의 에너지소모량에 비하여 V^II_r^It_r^I만큼 증가한다. 또한 BLE MCU는 이미지 MCU의 동작시간이 길어짐에 따라 식 5와 같이 기존 기법에 비하여 에너지소모량이 증가한다. 따라서 관심이미지가 촬영된 경우 식 6의 제안 기법의 평균 전력소모 P_{PI}는 기존 기법의 평균 전력소모 P_B보다 크다.

$$E_{PI}^I = V^I(I_i^I t_i^I + I_c^I t_c^I + I_r^I t_r^I + I_d^I t_t) \quad (4)$$

$$E_{PI}^B = V^B(I_s^B t_{IP}^B + I_t^B t_t + I_s^B (T - t_{IP}^B - t_t)) \quad (5)$$

$$P_{PI} = (E_{PI}^I + E_{PI}^B) / T \quad (6)$$

표 1. 기존 기법과 제안 기법의 비교 평가를 위한 저전력 영상인식 시스템의 동작 별 측정값
Table 1. Measurement of low-power image recognition system for evaluation of baseline and proposed methods

	Image MCU				BLE MCU		
	Initialize	Capture Image	Recognize Image	Deliver Image	TX Image	Idle	Sleep
Current Consumption	100mA	100mA	120mA	50mA	28mA	7mA	1mA
Time required	400ms	300ms	60ms	3000ms	3000ms	-	-
Power Consumption	300mW	300mW	360mW	150mW	84mW	21mW	3mW
Energy Consumption	120mWs	90mWs	21.6mWs	450mWs	252mWs	-	-

그러나 그림 2 (c)와 같이 관심이미지가 아닌 경우 제안 기법은 통신 과정의 생략이 가능하다. 따라서 이미지 MCU의 에너지소모량은 식 7과 같이 통신 과정의 생략으로 $V^I I_i^I t_i$ 만큼 감소한다. 그리고 식 8과 같이 BLE MCU의 에너지소모량은 통신 과정의 생략으로 대략 $V^B I_s^B t_s$ 만큼 감소한다. 따라서 식 9와 같이 관심이미지가 아닐 경우 제안 기법의 평균 전력소모 P_{PN} 는 기존 기법의 평균 전력소모 P_B 보다 개선시킬 수 있다.

$$E_{PN}^I = V^I (I_i^I t_i^I + I_c^I t_c^I + I_r^I t_r^I) \quad (7)$$

$$E_{PN}^B = V^B (I_i^B t_{iP}^B + I_s^B (T - t_{iP}^B)) \quad (8)$$

$$P_{PN} = (E_{PN}^I + E_{PN}^B) / T \quad (9)$$

전력소모 P_{PI} 는 영상인식 심층신경망을 위해 기존 기법의 전력소모 P_B 보다 크다. 그러나 전력소모 P_{PN} 은 P_B 에 비해 통신을 위한 전력소모를 절약할 수 있다. BLE 통신을 위한 전력소모 대비 영상인식 심층신경망을 위한 전력소모가 더 적을 경우, P_B 대비 P_{PI} 의 전력소모 증가에 비하여 P_{PN} 의 전력소모 감소량이 더 크다. 따라서 제안 기법은 식 10과 같이 관심이미지 비율 λ 에 따라 전력소모를 개선할 수 있다.

$$P_P = \lambda P_{PI} + (1 - \lambda) P_{PN} \quad (10)$$

IV. 실험

4.1 저전력 영상인식 시스템

MCU 장치는 500MHz 이하의 코어와 1MB이하의

SRAM을 가지는 제한적인 자원의 장치이다^[18]. MCU에서 영상인식 심층신경망을 활용하기 위해서는 심층신경망 모델의 크기를 줄이고 MCU를 위한 심층신경망 추론 가속 기술을 활용할 필요가 있다. 이를 위하여 기존 연구를 활용하여 CIFAR-10 데이터셋을 학습시킨 영상인식 심층신경망 모델에 대하여 8bit 양자화를 적용하여 모델을 경량화하고 MCU에서 심층신경망 가속을 위해 병렬연산을 지원하는 CMSIS-NN^[17] 라이브러리를 적용하였다. 본 연구에서는 이를 통해 이미지 MCU에서 이미지를 관심이미지인지 분류하는 영상인식 심층신경망을 구현하였다.

기존 기법과 제안 기법의 전력소모에 대한 실험을 위하여 저전력 영상인식 시스템의 프로토타입을 구현하였다. 장치의 이미지 MCU는 ST사의 STM32F765을 활용하여 구현하였고 BLE MCU는 nRF사의 nRF52840을 활용하여 구현하였다. 게이트웨이는 Raspberry Pi 3를 활용하여 장치와의 BLE 통신 기능을 구현하였다. 표 1은 오실로스코프와 전류 스코프를 활용하여 측정된 구현한 저전력 영상인식 시스템의 동작 별 전력소모이다. 제안한 저전력 통신 기법을 활용하여 구현한 영상인식 시스템의 이미지 MCU와 BLE MCU의 전력소모와 동작시간을 측정하였다. 이미지 MCU는 온디바이스 심층신경망을 활용하여 영상인식을 수행할 때 최대 전력소모인 360mW를 보였다. 그러나 MCU에서 심층신경망의 가속을 활용하여 소요시간을 낮춤으로써 에너지소모량은 21.6mWs로 매우 낮음을 확인하였다. 또한 이미지센서를 초기화하고 이미지를 촬영하는데 에너지소모량은 각각 120mWs와 90mWs임을 확인하였다.

클라우드 컴퓨팅 기반 영상인식 시스템의 경우 대부분의 전력소모를 촬영한 이미지를 클라우드 서버로 전송하는 동안 발생한다. BLE MCU에서 BLE 통신

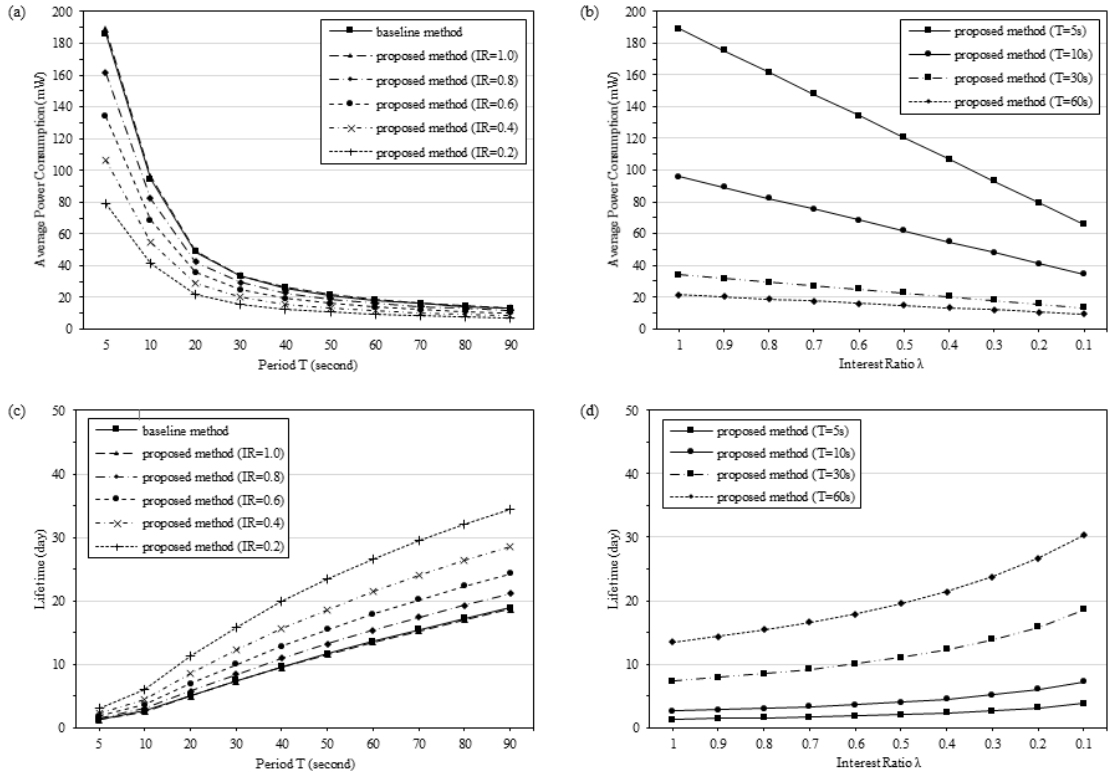


그림 3. (a) 주기에 따른 평균 전력소모 비교 (b) 관심이미지 비율에 따른 평균 전력소모 비교 (c) 주기에 따른 동작시간 비교 (d) 관심이미지 비율에 따른 동작시간 비교
 Fig. 3. (a) Comparison of average power consumption according to period (b) Comparison of average power consumption according to interest ratio (c) Comparison of lifetime according to period (d) Comparison of lifetime according to interest ratio

을 활용하여 이미지를 전달할 경우 전력소모는 84mW로 매우 낮다. 그러나 이미지의 크기로 인하여 통신 시간이 길어지므로 에너지소모량의 경우 252mWs로 매우 크다. 또한 이미지 MCU에서는 BLE MCU에 지속적으로 이미지를 전달하기 위하여 450mWs의 에너지를 소모하였다. 따라서 온디바이스 심층신경망을 활용하여 장치에서 영상인식을 수행하는 것보다 지속적으로 이미지를 전송하는 것이 에너지소모량이 더 큰 것을 확인할 수 있었다.

4.2 기존 기법 및 제안 기법의 전력소모 비교

그림 3 (a)와 그림 3 (b)는 각각 주기 T에 따른 기존 기법 및 제안 기법의 평균 전력소모와, 관심이미지 비율에 따른 제안 기법의 평균 전력소모를 나타낸다. 기존 기법의 평균 전력소모는 주기 T가 5초일 때 186.12mW, 10초일 때 94.56mW, 30초일 때 33.52mW, 60초일 때 18.26mW이다. 제안 기법의 평균 전력소모는 관심이미지 비율 가 0.98일 때 기존 기법과 동일하였다. 그리고 관심이미지 비율 가 0.2인

경우 제안 기법의 평균 전력소모는 주기 T가 5초일 때 79.416mW, 10초일 때 41.208mW, 30초일 때 15.736mW, 60초일 때 9.368mW이다. 따라서 제안 기법은 기존 기법에 대비 평균 전력소모를 대폭 개선하였다.

그림 3 (c)와 그림 3 (d)는 각각 주기 T에 따른 기존 기법 및 제안 기법의 동작시간과, 관심이미지 비율에 따른 제안 기법의 동작시간을 나타낸다. 장치의 배터리 용량은 AA건전지 두 개 용량인 6000mWh으로 설정하여 실험하였다. 기존 기법의 동작시간은 주기 T가 5초일 때 1.34일, 10초일 때 2.64일, 30초일 때 7.45일, 60초일 때 13.69일이다. 제안 기법의 동작시간은 관심이미지 비율 가 0.98일 때 기존 기법과 동일하였다. 그리고 관심이미지 비율 가 0.2인 경우 제안 기법의 동작시간은 주기 T가 5초일 때 3.14일, 10초일 때 6.06일, 30초일 때 15.88일, 60초일 때 26.68일이다. 따라서 제안 기법은 기존 기법에 대비 동작시간을 대폭 개선하였다.

4.3 비디오 데이터셋 실험

제안 기법은 기존 기법보다 관심이미지 비율에 따라 전력소모를 줄이고 동작시간을 향상시킨다. 본 절에서는 실제 비디오 데이터셋의 관심이미지 비율에 따라 기존 기법과 제안 기법을 비교 평가하였다. 비디오 데이터셋으로는 Caltech 비디오 데이터셋^[19]를 활용하였다. Caltech 비디오 데이터셋은 차량에 카메라를 설치하고 주행하며 보행자에 대한 비디오를 수집한 데이터셋이다. 해당 데이터셋의 관심이미지 비율은 대략 0.25이다. Caltech 비디오 데이터셋 환경에서 제안 기법의 평균 전력소모는 기존 기법에 비하여 주기 T가 5초일 때 46.34% 감소하였고, 주기 T가 60초일 때 54.4% 감소하였다. 그리고 동작시간은 주기 T가 5초일 때 2.89일, 주기 T가 60초일 때 25.15일을 보였다.

V. 결 론

본 논문에서는 저전력 영상인식 시스템을 위한 온디바이스 심층신경망 기반 저전력 통신 기법을 제안하였다. 주기적 스케줄링 및 다중 MCU의 기존 기법에 제안 기법을 결합하여 전력소모를 줄이고 동작시간을 향상시켰다. 그리고 제안 기법의 전력소모를 모델링하고 기존 기법과 전력소모와 동작시간을 비교하였다. 실제 차량카메라 데이터셋을 활용하여 제안 기법을 평가하여 한 달 이상 동작 가능한 저전력 영상인식 시스템을 구현할 수 있음을 확인하였다. 따라서 제안 기법을 활용할 경우 차량 카메라, 웨어러블 카메라, 감시 카메라 등 배터리를 기반으로한 다양한 영상인식 시스템의 동작시간을 개선할 수 있을 것으로 기대된다. 본 논문에서 제안한 온디바이스 심층신경망 기반 저전력 통신 기법은 심층신경망의 성능이 떨어질 경우 관심이미지의 손실률이 높아진다. 그러므로 MCU에서 심층신경망의 성능 향상에 대한 연구를 추후 진행할 계획이다.

References

[1] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, and F. E. Alsaadi, "A survey of deep neural network architectures and their applications," *Neurocomputing*, vol. 234, pp. 11-26, 2007.

[2] A. Botta, W. Donato, V. Persico, and A. Pescapé, "Integration of cloud computing and internet of things: A survey," *Future*

Generation Computer Syst., vol. 56, pp. 684-700, 2016.

[3] M. Aazam, I. Khan, A. A. Alsaffar, and E. Huh, "Cloud of things: Integrating internet of things and cloud computing and the issues involved," *11th IBCAST*, pp. 414-419, 2014.

[4] W. Chen, J. Wilson, S. Tyree, K. Weinberger, and Y. Chen, "Compressing neural networks with the hashing trick," *Int. Conf. Machine Learning*, 2015.

[5] W. Shi and S. Dustdar, "The promise of edge computing," *Computer*, vol. 49, no. 5, pp. 78-81, 2016.

[6] A. H. M. Amin, N. M. Ahmad, and A. M. M. Ali, "Decentralized face recognition scheme for distributed video surveillance in IoT-cloud infrastructure," *2016 IEEE Region TENSYP*, pp. 119-124, 2016.

[7] S. Chien, W. Chan, Y. Tseng, C. Lee, V. S. Somayazulu, and Y. Chen, "Distributed computing in IoT: System-on-a-chip for smart cameras as an example," *The 20th Asia and South Pacific Design Automation Conf.*, pp. 130-135, 2015.

[8] Y. Lee, J. Kim, and C. Kyung, "Energy-aware video encoding for image quality improvement in battery-operated surveillance camera," *IEEE Trans. VLSI Systems*, vol. 20, no. 2, pp. 310-318, 2012.

[9] K. Wolf, A. Schmidt, A. Bexheti, and M. Langheinrich, "Lifelogging: You're wearing a camera?," *IEEE Pervasive Computing*, vol. 13, no. 3, pp. 8-12, 2014.

[10] L. Yan, L. Zhong, and N. K. Jha, "Energy comparison and optimization of wireless body-area network technologies," *The ICST 2nd Int. Conf. Body Area Networks*, no. 8, 2007.

[11] R. Braojos, H. Mamaghanian, A. D. Junior, G. Ansaloni, D. Atienza, f. J. Rincon, and S. Murali, "Ultra-low power design of wearable cardiac monitoring systems," *IEEE/ACM Annu. Design Automation Conf.*, vol. 1, no. 1, pp. 140-145, 2014.

[12] C. Zhou, C. Tu, J. Tian, Y. Gao, and X. Ye,

“A low power miniaturized monitoring system of six human physiological parameters based on wearable body sensor network,” *Sensor Rev.*, vol. 35, no. 2, pp. 210-218, 2015.

- [13] B. Kim, S. Lee, J. Seo, and Y. Baek, “Design and implementation of personal low-power life logger applying hierarchical image capture algorithm based on multiple operating mode,” *J. KICS*, vol. 42, no. 12, pp. 2316-2325, 2017.
- [14] A. Howard, M. Zhu, B. Chen, and D. Kalenichenko, “MobileNets: Efficient convolutional neural networks for mobile vision applications,” *CVPR*, 2017.
- [15] F. Iandola, S. Han, M. Moskewicz, K. Ashraf, W. Dally, and K. Keutzer, “SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size,” arXiv preprint arXiv:1602.07360, 2016.
- [16] X. Zhang, X. Zhou, M. Lin, and J. Sun, “Shufflenet: An extremely efficient convolutional neural network for mobile devices,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2018.
- [17] L. Lai, “CMSIS-NN: Efficient neural network kernels for Arm Cortex-M CPU,” arXiv preprint arXiv:1801.06601, 2018.
- [18] Y. Zhang, N. Suda, L. Lai, and V. Chandra, “Hello edge: Keyword spotting on microcontrollers,” arXiv preprint arXiv:1711.07128, 2017.
- [19] P. Dollár, C. Wojek, B. Schiele, and P. Perona, “Pedestrian detection: A benchmark,” *CVPR*, Miami, FL, USA, Jun. 2009.

임재봉 (Jaebong Lim)



2016년 2월 : 부산대학교 컴퓨터공학과 공학학사
 2018년 2월 : 부산대학교 컴퓨터공학과 공학석사
 2018년 3월~현재 : 부산대학교 컴퓨터공학과 박사과정

<관심분야> 저전력 심층신경망, 임베디드 시스템
 [ORCID:0000-0002-1461-5693]

백윤주 (Yunju Baek)



1990년 2월 : 한국과학기술원 전산학과 공학학사
 1992년 2월 : 한국과학기술원 전산학과 공학석사
 1997년 2월 : 한국과학기술원 전산학과 공학박사
 1999~2002년 : NHN 연구소장

2003년~현재 : 부산대학교 컴퓨터공학과 정교수
 <관심분야> RFID, WSN, RTLS, 임베디드 시스템
 [ORCID:0000-0002-3873-2624]