Intel CPU기반 NVR 플랫폼에서 물리보안 구현을 위한 영상분석 구현에 관한 연구

송혁, 최인규, 유지상*

한국전자기술연구원, *광운대학교

{hsong, cig2982}@keti.re.kr, *jsyoo@kw.ac.kr

A Study on the image analysis for NVR security platform with Intel CPU

Hyok Song, In-Kyu Choi, Jisang Yoo*

Korea Electronics Technology Institute, *Kwangwoon Univ.

요 약

본 논문은 딥러닝 기술의 발전과 경량화로 인하여 엣지디바이스에서의 영상처리에 대한 요구가 증가하고 있음에 따라, 카메라 및 NVR 등의 엣지 디바이스에서의 영상 분석을 통한 다중디바이스 및 서버-디바이스 협력 시스템의 구현에 관한 논문이다. 본 논문에서는 Intel CPU 기반의 NVR 디바이스에서 영상보안 시스템을 구현하기 위하여 Intel CPU에서 지원하는 OpenVINO 툴킷을 활용하였으며 NVR 플랫폼상에 구현된 UI SW에 포함되어 상용 플랫폼상에서 동작함을 확인하였다. NVR 플랫폼에서의 실시간 동작을 위하여 보안시스템에서 필요한 행동인식의 정의, 행동인식 모델 개발, 영상개선 모델 개발, 영상개선 모델 최적화, OpenVINO 플랫폼을 통한 최적화를 구현하였다. 본 연구를 통한 연구 결과 NVR 플랫폼에서 영상인식 모듈이 80fps 이상의 속도로 실시간 구현됨을 확인하여 상용 플랫폼에서 활용 가능성을 확인하였다.

Ⅰ. 서 론

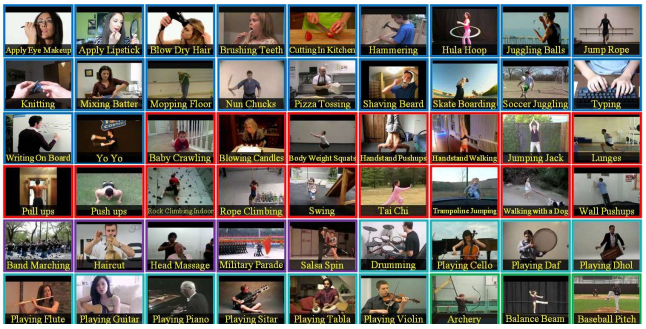
딥러닝 기술의 발전과 함께 영상 보안 기술의 수준이 향상되었으며 GPU 하드웨어 기술의 발전으로 딥러닝을 활용한 영상 분석 기술이 영상보안시스템에 사용되기 시작하였다. 과거 딥러닝 모델은 연산량이 매우 커 고성능 GPU에서 동작하였으나 최근 경량화된 딥러닝 모델의 개발 및 딥러닝 모델이 구동 가능한 엣지 디바이스용 칩셋이 출시됨에 따라 카메라 및 NVR 등의 엣지 디바이스에도 다양한 기능의 딥러닝 모델의 구동이 가능하다.

본 논문에서는 보안 시스템의 엣지 디바이스인 카메라 및 NVR에서 경량화된 영상 분석 모듈을 구동하기 위하여 개발된 NVR 플랫폼 상에서 행동인식, 영상개선 등의 딥러닝 모델을 구현하였으며 실시간 구동을 위하여 OpenVINO 툴킷을 사용하여 최적화 구현하였다[1].

본 논문에서는 연구개발 내용과 연구결과를 2장에서 보이며 3장에서는 향후 연구 방향을 제시하였다.

Ⅱ. 연구내용

1. 행동인식 기술



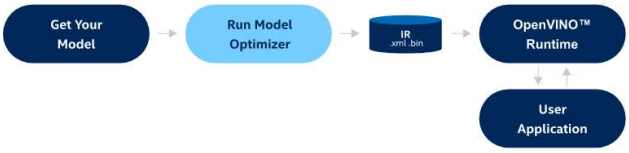
< 그림 1. UCF-101 데이터베이스 >

그림 1에서 보이는 바와 같이 UCF-101에서는 101종의 행동 카테고리를 정의하고 YouTube를 통하여 행동 데이터를 취득하여 활용하였으며 추후 50개의 데이터로 변경되었다[2]. Kinetic-400에서는 약 65만개의 영상 클립에서 데이터셋의 버전에 따라 400/600/700 종의 행동 클래스를 정의하였다[3]. 본 영상 데이터셋에는 단순한 행동 뿐 아니라 사람간 상호 인터랙션하는 데이터를 포함하고 있다. 그러나 영상보안 시스템에서 요구하는 행동의 정의는 기존 행동인식 클래수의 개수와 같이 다양한 행동 정의가 필요치 않아 실현장에서 요구하는 행동의 정의가 필요하다. 따라서 UCF-101 및 Kinetic-400 데이터에서 영상 보안에 적합한 데이터를 취합하여 그 데이터 위주의 학습을 진행하였다.

본 연구에서는 ResNet34 구조의 Encoder를 설계하여 프레임 단위의 빠른 특징을 추출하도록 설계하였으며 16 프레임 단위로 구성된 비디오 클립에 대한 실시간 행동 분류를 구현하였다[4].

< 그림 2. 동영상 데이터의 행동인신 결과 >

NPU 배포환경에서 최적의 동작을 위해 최적화 및 모델 경량화를 위한 양자화를 진행하였다. OpenVINO에서 행동인식 모델의 동작을 위한 과정으로 학습된 모델을 그림 3과 같이 중립모델로 변환하여 활용하였으며 그림 4와 같이 약 80fps의 결과를 보였다.



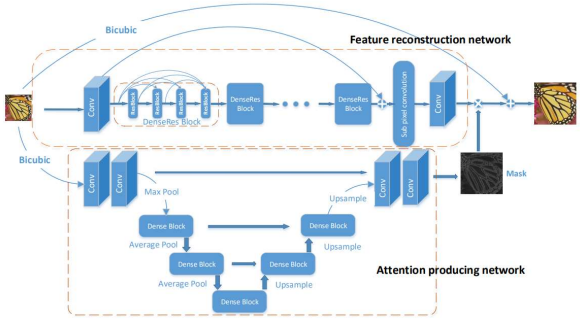
< 그림 3. OpenVINO 동작을 위한 중립모델 변환 >

```
[ INFO ] Metrics report:
[ INFO ]        Data total: 0.36ms (+/-: 1.28) 2758.92fps
[ INFO ]        Data   own: 0.35ms (+/-: 1.28) 2844.03fps
[ INFO ]        Encoder total: 12.36ms (+/-: 4.04) 80.88fps
[ INFO ]        Encoder   own: 12.35ms (+/-: 4.04) 80.98fps
[ INFO ]        Decoder total: 0.40ms (+/-: 0.47) 2515.02fps
[ INFO ]        Decoder   own: 0.39ms (+/-: 0.47) 2594.30fps
[ INFO ]        Render total: 33.36ms (+/-: 7.97) 29.97fps
[ INFO ]        Render   own: 33.10ms (+/-: 6.88) 30.21fps
```

< 그림 4. 행동인식 결과 >

2. 영상개선 기술

실현장 영상 보안 시스템에서 요구하는 주요 기능 중 하나로 상황 발생 시 이벤트 발생 구역에 대한 해상도를 높여 사용자가 원하는 데이터를 빨리 확인하도록 하는 기능을 요구하고 있으며 이는 SuperResolution 기술을 활용한다. 본 논문에서는 고속 영상 개선을 위하여 CNN 기반의 딥러닝 구조를 활용하였다. 본 논문에서는 그림 5와 같이 Attention-based Super Resolution 기술을 적용하였다[5].



< 그림 5. Attention-based Super Resolution 모델 구조 >

개발된 모델의 최적화하기 위하여 OpenVINO를 통하여 최적화 및 양자화를 하였고 중립모델을 통하여 NPU 환경에서 적용하여 그림 6과 같이 우수한 성능을 도출하였으며 그림 7과 같이 54.4msec의 속도를 보였다.

< 그림 6. NPU 환경에서 구현한 SR 결과 >

```
[ INFO ] The model /dev/shm/super_resolution/models/intel/single-image-super-resol
ution-1032/FP16-INT8/single-image-super-resolution-1032.xml is loaded to CPU
[ INFO ]        Device: CPU
[ INFO ]                Number of streams: 4
[ INFO ]                Number of threads: AUTO
[ INFO ]        Number of inference requests: 5
OpenCV: FFMPEG: tag 0x47504a4d/'MJPG' is not supported with codec id 8 and format
'image2 / image2 sequence'
[ INFO ] Metrics report:
[ INFO ]        Latency: 305.8 ms
[ INFO ]        FPS: 3.3
[ INFO ]        Decoding:       1.2 ms
[ INFO ]        Preprocessing:  2.9 ms
[ INFO ]        Inference:      54.4 ms
[ INFO ]        Postprocessing: 0.6 ms
[ INFO ]        Rendering:      nan ms
```

< 그림 7. SR 동작 결과 >

Ⅲ. 결론 및 향후 연구 계획

본 논문에서는 고성능 GPU 또는 CPU를 가진 플랫폼에서 동작하던 보안시스템을 위한 영상분석 기술을 실제 상용화 플랫폼인 NVR 및 카메라에서 실시간 동작을 구현하였다는데 의의가 있다. 구현된 행동인식 및 영상개선 딥러닝 모듈은 상용 UI SW에 포함되어 구동확인하였으며 그 동작 성능은 약 80fps의 행동인식 속도 및 18fps를 보였다. NVR 플랫폼은 Apollo Lake CPU를 가지고 있어 고성능 CPU가 아닌 저가형 모델에서 완성하였다.

현재 구현된 딥러닝 모델 및 최적화를 통하여 NVR에서 동작을 확인하여 이를 상용화하기 위한 시스템 통합 작업을 진행하고 있으며 본 작업을 통하여 향후 상용 제품으로 도출이 가능할 것으로 보인다. 향후 목표는 다양한 현장 요구를 취합하여 적절한 행동인식 클래스를 확장하고 영상 인식 기술등을 적용하여 기능을 확대하는데 있다.

ACKNOWLEDGMENT

This work was supported by the Technology development Program(S2977538) funded by the Ministry of SMEs and Startups(MSS, Korea).

참 고 문 헌

[1] https://docs.openvino.ai

[2] Soomro, Khurram, Amir Roshan Zamir, and Mubarak Shah. "UCF101: A dataset of 101 human actions classes from videos in the wild." arXiv preprint arXiv:1212.0402 (2012).

[3] Smaira, Lucas, et al. "A short note on the kinetics-700-2020 human action dataset." arXiv preprint arXiv:2010.10864 (2020).

[4] Koonce, Brett. "ResNet 34." Convolutional Neural Networks with Swift for Tensorflow. Apress, Berkeley, CA, 2021. 51-61.

[5] Pesavento, Marco, Marco Volino, and Adrian Hilton. "Attention-based multi-reference learning for image super-resolution." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021.