Journal of KIIT. Vol. 18, No. 9, pp. 57-67, Sep. 30, 2020. pISSN 1598-8619, eISSN 2093-7571
http://dx.doi.org/10.14801/jkiit.2020.18.9.57

고주파 통과 필터와 합성곱 신경망을 이용한 동영상 촬영 장치 판별 알고리즘

김동현*, 이해연**

Video Capturing Device Identification Algorithm based on High-Pass Filter and Convolutional Neural Network

Dong-Hyun Kim*, Hae-Yeoun Lee**

This work was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2017R1D1A1B03030432, 2020R1F1A1057742).

요 약

IT 기술의 발전으로 누구나 손쉽게 고성능의 동영상 촬영기기를 손쉽게 접할 수 있고 동영상을 생산하고 있다. 그러나 이들 동영상의 불법적인 활용으로 인하여 많은 사회적인 문제들을 야기하고 있어서 동영상 포렌식 기술에 대한 필요성이 높다. 본 논문에서는 동영상 포렌식 기술 중에 하나로서 고주파 필터와 합성곱 신경망을 이용하여 동영상 촬영 장치를 판별하기 알고리즘에 대하여 제안한다. 기존의 많은 연구들이 영상 촬영 장치에 초점을 맞추고 있으나 동영상의 경우 프레임 특성, 압축 방법, 압축률 및 방대한 용량 등에 있어서 영상과 차이가 있어서 영상을 대상으로 한 기술의 직접적인 적용이 불가능하여 동영상에 최적화된 알고리즘을 개발하였다. 제안한 알고리즘의 성능을 분석하기 위하여 동일한 브랜드 및 모델을 포함한 총 31개의 다양한 촬영 장치로 수집한 동영상 데이터를 사용하였고 91.3%의 판별 정확도를 달성하였다. 또한, 동일 브랜드의 모델에 대해서도 91.1% 정확도를 달성하여 기존 연구들와 다르게 더 높은 정확도를 갖고 있는 것을 보였다.

Abstract

Advances in IT technology make it easy for anyone to access high performance but low-cost video capturing devices and produce videos. However, these videos cause many social problems due to illegal use, and thus video forensic technology is required. In this paper, as one of the video forensic technologies, we propose an algorithm for identifying a video capturing device using a high frequency filter and a convolutional neural network. Many previous studies have focused on image capturing devices not video capturing devices. In videos, there are differences from images in frame characteristics, compression method, compression rate, and huge capacity. Therefore, it is impossible to directly apply the technology targeting the images and hence we developed an optimized algorithm for the videos. In order to analyze the performance of the proposed algorithm, videos captured by 31 various capturing devices including the same brand and model were used and we achieved an 91.3% accuracy. Also, we showed that the proposed algorithm can identify the models of the same brand with 91.1% accuracy higher than the previous studies.

Keywords
video capturing device identification, high-pass filter, convolutional neural network, video forensics

* 금오공과대학교 소프트웨어공학과 석사과정
- ORCID: https://orcid.org/0000-0002-0693-431X
** 금오공과대학교 컴퓨터소프트웨어공학과 교수(교신저자)
- ORCID: https://orcid.org/0000-0002-6081-1492

· Received: Jun. 17, 2020, Revised: Jul. 13, 2020, Accepted: Jul. 16, 2020
· Corresponding Author: Hae-Yeoun Lee
Dept. of Computer Software Engineering, Kumoh National Institute of Technology, Korea
Tel.: +82-54-458-7548, Email: haeyeoun.lee@kumoh.ac.kr

Ⅰ. 서 론

정보 통신 기술의 급속한 발전으로 동영상 촬영 장치는 비디오 카메라에서 스마트 폰, 디지털 카메라 등으로 다변화되었으며, 고품질과 고성능을 가지고 있으나 저렴한 가격으로 누구나 손쉽게 접근이 가능해졌다. 따라서 이들 장치를 이용하여 동영상을 촬영하고 유튜브, 인스타그램, 페이스북 등과 같은 SNS 및 유사 플랫폼을 통하여 배포하면서 정보를 공유하고 다양한 산업 분야의 창출이 이루어지고 있다. 그러나 이와 같은 이점에도 불구하고 불법적 목적을 가지고 있는 사용자가 동영상을 불법적으로 활용함에 따라서 다양한 사회적 문제 및 범죄가 발생하고 있다. 표 1에는 대검찰청에서 발표한 10년간 카메라 등을 이용한 성범죄 및 통신매체를 이용한 성범죄 통계를 보여주고 있으며 증가하고 있는 추세를 확인할 수 있다[1].

표 1. 10년간 카메라 / 통신매체 이용한 성범죄 현황
Table 1. Sexual crimes using camera or internet for last 10 years

Year	Sexual crimes using camera	Sexual crimes using internet
2009	834	761
2010	1,153	1,031
2011	1,565	911
2012	2,462	917
2013	4,903	1,416
2014	6,735	1,254
2015	7,730	1,139
2016	5,249	1,115
2017	6,615	1,265
2018	6,085	1,378

이와 같은 촬영 장치를 이용한 범죄의 수사나 예방을 위하여 포렌식 기술에 대한 필요성이 높아지고 있으며 미국 및 유럽 등 선진국을 중심으로 활발히 연구가 진행되었고 성과들이 나타나고 있다. 촬영 장치 판별 포렌식 기술은 초기에 센서의 불완전성에 기인하는 장치별 고유의 특징을 활용하여 판별하는 기술들이 연구되었고, 최근에는 딥러닝을 접목한 연구들이 진행되고 있다. 그러나 이들 대부분은 영상 촬영 장치에 대한 연구가 다수이며 동영상 촬영 장치에 대한 연구는 상대적으로 부족하다.

본 논문에서는 동영상을 대상으로, 최근 관심을 많이 받고 있는 딥러닝 기술을 적용한 고주파 필터와 합성곱 신경망을 이용한 동영상 촬영 장치 판별 포렌식 알고리즘을 제안한다. 동영상은 영상과 다르게 I 프레임(Intra frame), P 프레임(Predictive frame), B 프레임(Bi-directional predictive frame) 특성을 갖고 있고, 압축 방법과 압축률이 다르며, 방대한 용량을 가지고 있기 때문에 영상을 대상으로 하는 알고리즘의 직접적 적용이 어렵다. 따라서, 동영상 프레임별 특성을 분석하여 96.2%의 정확도를 갖는 P 프레임을 활용하고, 합성곱 계층 수에 따른 성능을 분석하여 97.9%의 정확도를 갖는 4계층 모델을 선택하여 동영상에 적합한 알고리즘을 개발하였다.

동일한 브랜드 및 모델을 포함한 총 31개 다양한 촬영 장치로 수집한 방대한 양의 동영상을 사용하여 실험을 수행하였고, 제안한 알고리즘이 91.3%의 판별 정확도를 갖고 있음을 보였다. 또한, 제안한 알고리즘이 기존 연구들보다 더 높은 정확도 91.1%로 동일 모델에 대한 판별이 가능한 것도 보였다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구를 제시하고, 3장에서는 제안하는 알고리즘을 설명한다. 실험 및 분석 결과는 4장에서 제시하고 5장에서 결론을 기술한다.

Ⅱ. 관련 연구

2.1 특징 기반 촬영 장치 판별 연구

촬영 장치 판별을 위한 기존의 연구들은 동영상이 아닌 영상 촬영 장치를 대상으로 하고 있으며, 촬영 센서의 불완전성에 기인하는 고유한 잡음 특징을 활용하는 연구가 대부분이며, 장치의 색상 필터 행렬에 대한 보간 특성, 암전류 특성 등을 활용하는 연구들도 진행되었다.

Lukas 등 및 Chen 등은 영상 촬영 과정에서 물리적 센서의 불완전성에 의하여 의도하지 않게 발생하는 센서 고유의 패턴 잡음(PRNU, Photo Response Non-Uniformity)를 이용하여 촬영 장치를 판별하기 위한 연구를 수행하였다[2][3].

Bayram 등은 컬러 영상을 생성하기 위해서는 색상 필터 행렬(Color filter array)를 통과한 신호에 대

하여 보간의 과정을 거치는데 이때 발생하는 화소 사이의 상관 관계를 이용하여 촬영 장치 판별하기 위한 연구를 수행하였다[4].

Kuroki 등은 캠코더에 포함된 CCD(Charged Coupled Device) 센서에 흐르는 암전류(Dark current)의 비균일성을 이용하여 촬영 장치 판별하기 위한 연구를 수행하였다[5].

이와 같은 연구들은 사람이 직접 고안한 특징을 활용하여 판별하기 위한 기술들로서, 새로운 기기들이 빠르게 나타나는 상황에서 사람이 직접 특징을 고안하는 것이 용이하지 않다.

2.2 딥러닝 기반 촬영 장치 판별 연구

최근에는 딥러닝 기술을 도입하여 촬영 장치를 판별하기 위한 연구들이 진행되고 있다. 이를 통하여 사람이 장치들을 판별하기 위한 특징을 직접 고안하지 않고, 최적의 특징을 인공 지능을 통하여 학습을 하도록 하고 있다.

Tuama 등은 고주파 통과 필터와 3개의 합성곱 계층과 3개의 전연결 계층을 갖는 딥러닝 모델을 이용한 카메라 모델 판별 연구를 수행하였다[6]. Dresden 데이터베이스에서 27개 장치 및 추가적인 6개 장치로 촬영한 영상을 이용하여 성능을 분석하였고, 91.9% 정확도를 달성하였다.

Wang 등은 밝기 변화에 무관하도록 LBP(Local Binary Pattern) 코딩을 이용하고 3개의 합성곱 계층과 3개의 전연결 계층을 갖는 딥러닝 모델을 이용한 모델 판별 연구를 수행하였고 Dresden 데이터베이스의 12개 장치를 활용하여 98.78%의 판별 정확도를 달성하였다[7].

기존의 합성곱을 활용한 모델의 경우 계층을 통과하는 과정에서 해상도가 줄어드는 문제가 발생하므로 Kamal 등은 Dense 합성곱 신경망을 이용한 모델 판별 연구를 수행하였다[8]. 1개의 합성곱 계층과 3개의 Dense 블록 계층으로 네트워크를 구성하였고 Dresden 데이터베이스의 19개 카메라 모델에 대하여 99% 이상의 정확도를 달성하였다.

합성곱 신경망에서 계층이 얕은 경우 성능이 저하되는 문제를 해결하기 위하여 Chen 등은 잔차 매핑을 활용하는 잔차 뉴럴 네트워크(ResNet, Residual Neural Network)를 이용한 카메라 모델 판별 연구를 수행하였다[9]. 5개의 합성곱 계층과 1개의 전연결 계층으로 구성하였고, 브랜드 판별에서는 13개 카메라 대상으로 99.1% 정확도를 보였고, 모델 판별에서는 27개 카메라 대상으로는 94.7% 정확도를 보였으며, 장치 판별에서는 73개의 카메라를 대상으로 45.8% 정확도를 보였다.

이와 같은 딥러닝을 활용한 연구들의 경우도 대부분 영상 촬영 장치를 대상으로 연구를 진행하고 있고 동영상 촬영 장치를 대상으로 하는 연구는 많지 않다. 동영상은 영상과는 다른 프레임 특성을 갖고 있고, 압축 방법과 압축률이 다르며, 방대한 용량을 가지고 있어서 연구를 수행하는 것이 용이하지는 않다.

2.3 영상과 동영상 압축에 대한 비교

일반적으로 영상은 인접한 픽셀 사이의 밝기값 차이가 평균적으로 크지 않고 유사한 공간적 중복성(Spatial redundancy) 특징을 갖는다. 따라서 영상에 대한 압축은 이와 같은 공간적 중복성을 제거하는데 주로 초점을 맞추고 있다.

그러나 동영상은 다수의 프레임들로 구성이 되어 있으며, 영상처럼 공간적 중복성 뿐만이 아니라 이전 프레임의 픽셀 밝기값이 다음 프레임의 픽셀 밝기값과 유사한 시간적 중복성(Temporal redundancy)을 갖고 있다. 따라서 압축을 수행할 때 각 프레임 독립으로 처리하지 않고 인접 프레임 사이의 상관 관계를 활용하여 압축률을 높인다.

동영상은 그림 1과 같이 GoP(Group of Pictures)로 부르는 프레임들의 집합으로 구성된다.

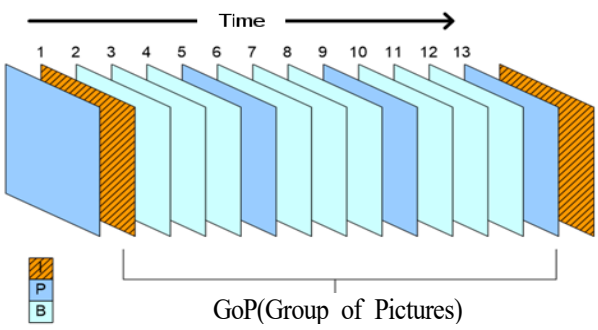


그림 1. 동영상 프레임의 그룹 및 종류
Fig. 1. Group and types of video frames

시간적 중복성은 시간에 따라 나열된 프레임의 집합이란 점에서 연속된 프레임 사이에 갑작스럽고 큰 변화가 없는 한 실제로는 프레임에 속한 픽셀들이 조금씩 변화할 뿐이다. 이 경우 실제로 배경과 같은 대부분의 정적인 요소들은 연속된 프레임 사이에 거의 불변이며, 물체의 움직임만 차이가 난다. 동영상은 이러한 시간적 중복성을 제거하기 위하여 연속된 프레임 사이에 변하지 않는 배경과 같은 정적인 요소들에 의한 영향을 제거하는 기법들을 활용하며 이것이 영상 압축 방법들과 가장 큰 차이점이다. 또한 동영상은 일반 영상에 비해 많은 정보를 연속적으로 표현해야 하므로 많은 저장 공간을 요구하므로 높은 압축률로 압축을 한다.

GoP는 인트라 프레임과 비인트라 프레임으로 구성되어 있다. 인트라 프레임(I 프레임)은 영상과 마찬가지로 DCT를 이용하여 공간적 중복성만 제거한 프레임으로 JPEG와 유사한 압축방식으로 압축된 프레임이다. 하지만 JPEG의 압축과 유사할 뿐, 실제 압축 방법은 상이하다. 일반적으로 I 프레임은 키 프레임으로 불리며 연속되는 화면의 기준이 된다. 비인트라 프레임은 예측 프레임(P 프레임)과 양방향 예측 프레임(B 프레임)으로 구성되어 있다. P 프레임은 앞에 나타난 I 프레임 또는 다른 P 프레임을 참조하여 움직임 보상 처리 후에 움직임 벡터와 예측 오차를 인코딩하여 처리한 프레임이다. B 프레임은 앞과 뒤의 I 프레임과 P 프레임의 2개의 프레임을 이용하여 움직임 보상을 거친 프레임이다. 특히 B 프레임의 경우 이전의 프레임과 이후의 프레임간의 차이를 평가하여 구성하므로 시간적인 연속성에 보장되지 않는다.

III. 고주파 통과 필터와 합성곱 신경망을 이용한 동영상 촬영 장치 판별 알고리즘

동영상 촬영 장치 판별을 위한 제안하는 알고리즘의 구조는 그림 2와 같이 학습의 과정과 판별의 과정으로 구성되어 있다.

학습의 과정을 위해서는 입력 데이터로 동영상과 장치에 대한 정보가 필요하며, 판별의 과정에서는 입력 데이터로 미지의 동영상을 입력 받는다.

동영상은 다양한 종류의 프레임들로 구성이 되어 있으나, 제안한 알고리즘에서는 학습과 판별의 과정에서 P 프레임만 추출하여 학습과 판별에 사용하였다.

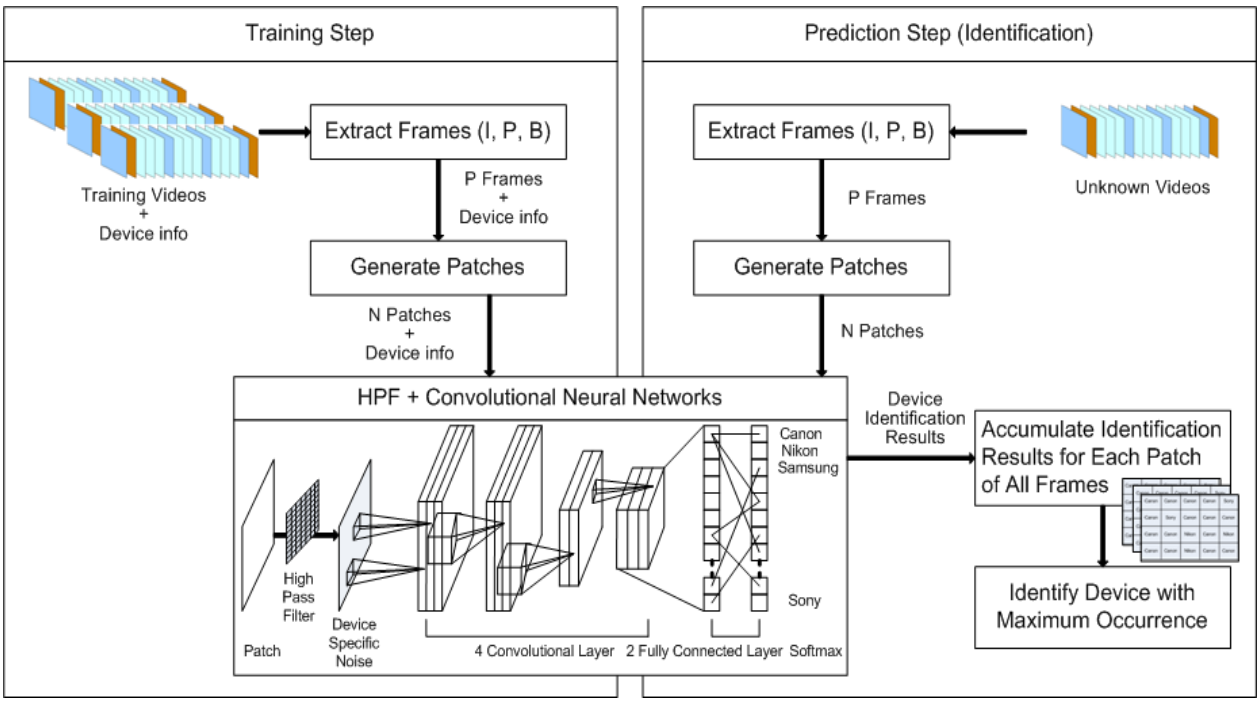


그림 2. 제안하는 동영상 촬영 장치 판별 알고리즘 구조

Fig. 2. Structure of the proposed video capturing device identification algorithm

또한, 동영상 프레임 해상도가 F-HD(1920×1080), Q-HD(3840×2160) 등으로 매우 크기 때문에 딥러닝 모델을 실행하기 위한 하드웨어 메모리 등의 제약으로 인하여 작은 패치로 분할하여 학습과 판별에 활용하였다. P 프레임만을 대상으로 처리하도록 설계한 이유에 대해서는 3.1절에서 설명한다.

동영상 촬영 장치 판별을 위해서는 고주파 통과 필터와 합성곱 신경망을 이용한 모델을 활용하였다. 고주파 통과 필터는 센서로 인한 영상 내의 노이즈를 효율적으로 추출할 수 있고, 이를 장치 인식에 활용하는 것이 좀 더 높은 정확도를 달성할 수 있어서 적용하였다[10]. 합성곱 계층을 4계층으로 정한 이유는 3.2절에서 기술하고, 구체적인 모델 구조는 3.3절에서 설명한다.

학습을 수행한 후에 미지의 동영상에 대한 판별을 수행할 때에 동영상은 다수의 프레임(M개)으로 구성이 되어 있고, 각 프레임을 여러 개의 패치(N개)로 분할하여 장치에 대한 판별을 수행하였으므로, 최종적인 동영상 촬영 장치에 대한 판별은 M×N개의 판별 결과 중에서 최대 빈도수를 갖는 장치를 판별한 결과로 결정한다.

3.1 최적 동영상 프레임 종류 설정

동영상은 I 프레임, P 프레임, B 프레임으로 구성이 되어 있으며, 각 프레임 유형마다 압축 방식이 다르다. I 프레임은 공간적 중복성만을 제거하고, P 프레임과 B 프레임은 시간적 중복성도 제거한다. 따라서 I 프레임은 정보의 손실이 가장 작은 특성이 있고, P 프레임과 B 프레임은 다른 프레임의 정보를 고려하여 차이가 있는 영역에 대해서 움직임 보상 등 처리를 통하여 압축을 수행하므로 압축률이 높으나 정보 손실이 상대적으로 크다. 특히 B 프레임은 시간적으로 과거와 미래에 존재하는 프레임을 활용하여 압축을 수행하는데, 실시간 촬영을 하며 압축하는 경우 하드웨어 성능에 대한 제약과 미래의 프레임 정보를 활용할 수 없어서 압축된 동영상에 보통은 포함되지 않는다.

따라서 동영상 프레임 종류에 따른 3.2절에서 설명하는 고주파 통과 필터와 합성곱 신경망으로 구성된 딥러닝 모델의 성능 분석을 위하여 5개 디지털 카메라 장치로 동영상을 촬영하였고, 각 동영상에서 프레임을 종류별로 추출하였다. 프레임의 해상도가 하드웨어 성능에 비하여 크기 때문에 256×256 픽셀의 패치로 분할을 수행한 후에 학습과 판별을 수행하였고, 정확도에 대한 분석을 수행하여 표 2에 정리하였다. 다수의 장치를 사용하여 실험을 수행하는 것은 계산 자원과 시간 측면에서 비효율적으로, 5개 장치로 제한을 하여 실험하고 분석하였다.

표 2. 동영상 프레임 종류에 따른 판별 성능 분석
Table 2. Identification accuracy analysis depending on video frame types

		Training data		
		I frame	P frame	I+P frame
Testing data	I frame	78.3%	58.0%	60.2%
	P frame	84.4%	**96.2%**	95.5%
	I+P frame	89.4%	96.1%	95.7%

분석한 결과에 따르면 P 프레임으로 딥러닝 모델의 학습을 수행한 후에 P 프레임으로 판별을 수행하는 것이 96.2%로 매우 높은 정확도를 갖는 것으로 나타났다.

동영상 촬영 장치를 판별하는 측면에서 I 프레임이 압축으로 인한 정보의 손실이 작아서 판별의 정확도가 높을 것으로 예상되지만 프레임 개수가 적어서 정확도가 낮은 것으로 나타났다.

P 프레임만 사용하는 경우 압축 과정에서 움직임 보상 등으로 인하여 위치 정보에 대한 동기화가 어긋나기 때문에 정보의 손실이 I 프레임보다 많이 발생하지만 다수의 프레임이 존재하여 오히려 유리한 것으로 나타났다.

I 프레임과 P 프레임을 동시에 활용하는 경우도 높은 성능을 갖고 있는 것으로 나타났지만, P 프레임만을 사용하는 것보다는, 오히려 I 프레임이 학습을 방해하는 효과를 나타내었다.

3.2 합성곱 계층수에 따른 판별 정확도

영상을 대상으로 하는 카메라 모델 판별을 위한 기존 연구들에서는 고주파 통과 필터와 3개 합성곱 계층으로 구성된 합성곱 신경망을 일반적으로 활용하였다. 그러나 최적의 합성곱 계층의 수나 파라미

터 등에 대한 분석은 제시되지 않았다.

본 논문에서 제안하는 동영상 촬영 장치 판별 알고리즘에서도 범용적으로 활용되는 고주파 통과 필터와 합성곱 신경망 모델을 설계하여 활용하였다. 다만, 동영상 촬영 장치 판별에 있어서 최적의 성능이 나타나도록 모델에 대한 계층의 구성, 처리 함수와 파라미터에 대한 최적화를 수행하였다.

최적의 합성곱 계층 수를 결정하기 위하여 12개 장치로 촬영한 256×256 해상도 컬러 영상에 대하여 2개~5개 다양한 컨볼루션 계층 수를 설정하여 판별 정확도에 대한 실험을 수행하였고, 그 결과를 표 3과 그림 3에 정리하였다. 그림 3의 X축은 epoch 및 Y축은 정확도를 나타낸다.

다수의 장치 데이터를 활용하는 것은 높은 계산 시간과 자원을 필요로 하여 효율성을 위하여 12개 장치로 제한하여 실험하였다. 공통적으로 입력 영상에 고주파 통과 필터를 적용하였으며, 전연결 계층은 2개로 하였고, 맥스 풀링(Max Pooling) 기법을 적용하였다. 기존 연구들에서는 3계층을 많이 사용하지만 표 3과 그림 3에 나타난 것과 같이 4계층을 활용하는 것이 수렴 속도도 빠르며 정확도가 더 높은 것으로 나타났다. 오히려 계층의 수가 더 많아지는 경우 정확도가 더 저하되는 것으로 판단된다.

표 3. 합성곱 계층의 설정에 따른 판별 성능
Table 3. Identification accuracy depending on the setting of convolutional layers

	2 layers	3 layers	4 layers	5 layers
Convolutional layer output	64/32	64/64/32	64/64/ 128/32	64/64/ 128/128/32
Accuracy	94.0%	97.1%	97.9%	95.4%

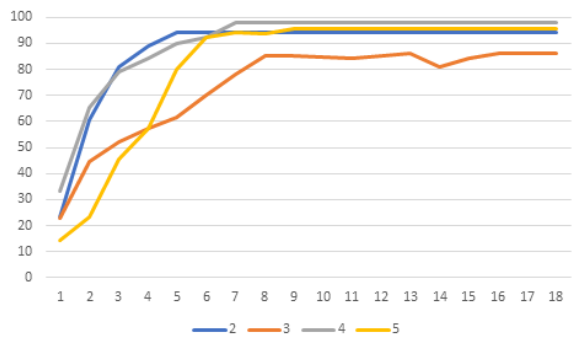


그림 3. 합성곱 계층 수에 따른 판별 정확도
Fig. 3. Identification accuracy depending on the number of convolutional layers

3.3 고주파 통과 필터와 합성곱 신경망 모델

그림 2의 동영상 촬영 장치 판별 알고리즘에 적용된 고주파 통과 필터와 합성곱 신경망으로 설계된 동영상 촬영 장치 판별 모델에 대하여 계층 구성, 처리 함수, 파라미터 설정 등을 구체적으로 도시하면 그림 4와 같다.

각 프레임에서 추출한 패치들을 입력 받은 후에, 2절에서 설명한 특징 기반 장치 판별의 연구들과 유사하게 입력된 패치에 존재하는 하드웨어 촬영 센서의 고유한 흔적 즉 특징을 추출하기 위하여 고주파 통과 필터를 적용한다. 고주파 통과 필터를 위해서는 다음과 같이 범용적으로 활용되는 5×5 크기의 마스크를 사용하였다.

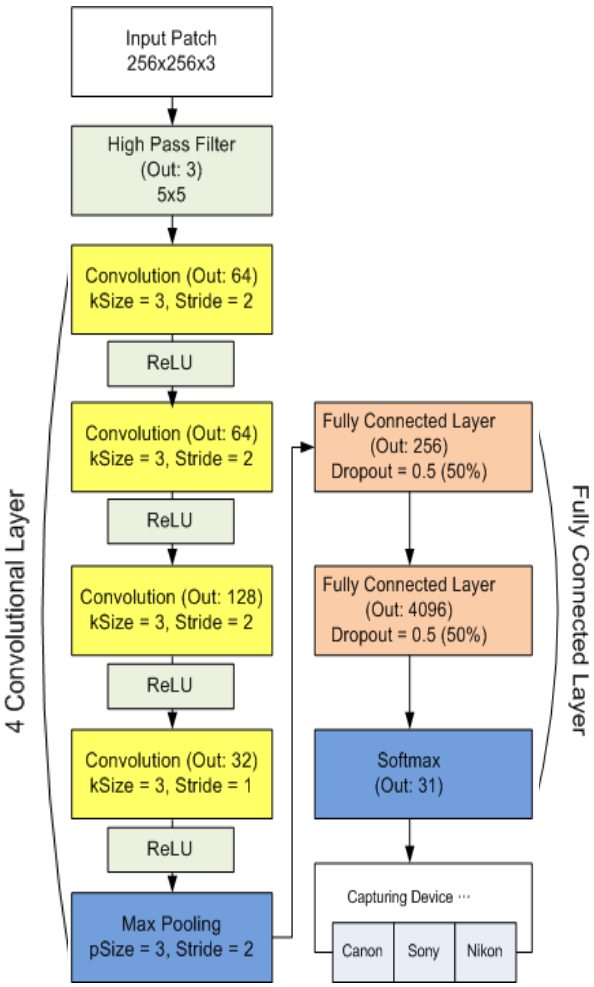


그림 4. 고주파 통과 필터와 합성곱 신경망 기반의 동영상 촬영 장치 판별 모델
Fig. 4. Video capturing device identification model based on high-pass filter and convolutional neural network

$$HPF = \frac{1}{12}\begin{pmatrix} -1 & 2 & -2 & 2 & -1 \\ 2 & -6 & 8 & -6 & 2 \\ -2 & 8 & -12 & 8 & -2 \\ 2 & -6 & 8 & -6 & 2 \\ -1 & 2 & -2 & 2 & 1 \end{pmatrix} \quad (1)$$

컨볼루션 계층은 3.2절의 실험과 같이 4계층이 최적으로 분석되어 4개로 구성하였으며 첫 번째 및 두 번째 컨볼루션 계층은 커널 크기가 3이고 stride가 2이며 출력은 64개로 설정하였다. 세 번째 컨볼루션 계층은 커널 크기가 3이고 stride를 2로 하고 출력은 128로 설정하였다. 마지막 네 번째 컨볼루션 계층에서는 커널 크기가 3이고, stride를 1로 하였으며 출력은 32로 설정하였다.

모든 계층의 활성화 함수로는 0 이하의 값을 0으로 설정하고, 0 초과의 값을 그대로 유지하는 ReLU(Rectified Linear Units)를 이용하였다[7].

컨볼루션 계층의 마지막에는 출력의 크기를 줄이기 위해 맥스 풀링 처리를 수행하며 파라미터값으로 커널 크기는 3, stride는 2로 하였다. 맥스 풀링 과정을 통하여 2×2 값 중에 최대값 1개를 선택하여 남김으로써 크기를 줄일 수 있다[11].

컨볼루션 계층에 대한 처리를 수행하고 난 후에 전연결 계층은 2개로 구성하였고, 두 계층 모두 Drop-out을 0.5로 설정하였다. 첫 번째 계층의 경우 256개의 노드로 구성하였고, 두 번째 계층의 경우 4096개의 노드로 구성하였다. 마지막으로는 소프트맥스(Softmax) 과정을 통해 최대값을 검색하여 각 촬영 장치에 대한 판별을 수행한다.

Ⅳ. 실험 및 성능 분석

4.1 실험 환경과 데이터

동영상 촬영 장치 판별을 위한 알고리즘의 실험과 분석을 위하여 CPU는 Intel i7-7700, 그래픽카드는 nVidia TitanXP(VRAM 12GB)를 사용하였다. 메모리는 16GB로 구성하였고, Windows 10 환경에서 Tensorflow GPU(1.15.0 버전) 프레임워크를 사용하여 알고리즘을 구현하였다.

실험을 위해서 총 31개의 동영상 촬영 장치를 사용하여 동영상 데이터를 확보하였고, 각 장치 모델, 동영상 해상도, 학습과 판별에 사용한 동영상 개수 및 총 I 프레임과 총 P 프레임의 개수를 표 4에 정리하였다.

표 4. 촬영 장치, 동영상 해상도 및 프레임 수
Table 4. Video capturing device, video resolution and frame numbers

Device	Resolution	Training video			Testing video		
		#	I Frame	P Frame	#	I Frame	P Frame
Canon 650D	1920x1080	11	401	4354	10	358	3903
Canon EOS 500D	1920x1080	10	416	3697	10	404	3577
Canon EOS M	1280x720	10	761	8326	10	756	8259
Canon IXUS 160	1280x720	12	330	3588	10	336	3627
Nikon Coolpix S33	1920x1080	10	151	4379	10	158	4582
Nikon Coolpix S100	1920x1080	6	65	1885	4	51	1479
Panasonic DMC SZ1	1280x720	7	163	2282	6	206	2884
SAMSUNG WB35F	1280x720	11	516	1857	10	627	1645
GoPro	1280x720	9	1494	10426	12	2247	15688
Lumix DWC LX100	3840x2160	12	346	1384	12	400	1600
G Pro2	3840x2160	7	109	3124	8	123	3558
G2	1920x1080	8	120	3402	7	102	2868
G3 Cat6	1920x1080	10	156	4417	10	155	4335
G4	1920x1080	8	118	3419	7	104	2964
Galaxy Grand Max	1920x1080	8	131	3668	7	113	3210
Galaxy Note2 A	1920x1080	7	79	2217	8	90	2536
Galaxy Note2 B	1920x1080	8	101	2786	7	83	2324
Galaxy Note3 A	1920x1080	7	77	2127	8	91	2520
Galaxy Note3 B	1920x1080	8	131	3678	7	116	3261
Galaxy Note3 C	1920x1080	8	129	3629	7	114	3200
Galaxy S4 LTE	1920x1080	8	89	2497	7	80	2208
Galaxy S5	1920x1080	8	89	2451	7	76	2101
iPhone 5S A	1920x1080	7	107	3015	8	124	3454
iPhone 5S B	1280x720	9	70	968	6	65	906
iPhone 6 A	1920x1080	8	127	3545	7	106	3041
iPhone 6 B	1920x1080	16	205	5790	14	177	5046
iPhone 6 C	1920x1080	8	125	3512	7	114	3218
iPhone 6Plus	1920x1080	8	122	3408	7	109	3032
Vega Secret Note	1920x1080	8	94	2561	7	81	2255
Vu3	1440x1080	7	108	3024	8	106	2950
Galaxy Alpha	1920x1080	7	69	1897	5	55	1571
Sum		271	6,999	107,313	253	7,727	107,802

고품질 DLSR과 미러리스 카메라, 범용 디지털 카메라 및 스마트폰 카메라 등 다양한 장치를 이용하여 데이터를 수집하였고, FHD(1920×1080) 및 QHD(3840×2160) 등 다양한 해상도의 데이터를 갖고 있다. 학습과 판별에 사용한 동영상 개수는 유사하지만 녹화 시간 등의 차이로 인하여 프레임 수에는 차이가 있으나, P 프레임이 I 프레임보다 10배 이상 많은 것을 확인할 수 있다.

특히, 동일한 모델에 대한 판별 성능 분석을 위해 동일한 모델이지만 다른 장치로는 Galaxy Note2, Galaxy Note3, iPhone 5S, 및 iPhone 6가 포함되어 있음을 확인할 수 있다.

4.2 동영상 촬영 장치 판별 정확도 분석

각 장치로 촬영한 동영상에서 P 프레임을 추출하고, 각 프레임을 256×256 패치로 분할하여 딥러닝 모델에 대한 학습을 수행한 후에 패치 별로 판별한 결과와 최대 빈도수를 기준으로 동영상에 대한 판별을 수행한 정확도 결과를 표 5에 요약하였다.

패치 판별 정확도의 경우 특정 장치 모델로 촬영한 동영상의 전체 패치 중에서 해당 장치 모델로 촬영한 것으로 올바르게 판별한 패치의 비율을 의미하며, 총 31개 동영상 촬영 장치를 대상으로 69.7%의 정확도를 보였다. G2, Galaxy S4 LTE, iPhone 6 C 모델의 경우 패치별 판별 정확도가 40% 초반인 것을 확인할 수 있다.

동영상 판별 정확도의 경우 특정 장치로 촬영한 동영상에 대하여 해당 장치 모델로 촬영한 것으로 올바르게 판별한 동영상의 비율을 의미하며 총 31개 동영상 촬영 장치를 대상으로 91.3%의 정확도를 보였다. 학습 동영상이 아닌 판별 동영상만을 대상으로 정확도를 계산하였다. 대부분 장치에서 100%로 올바르게 판별하고 있는 것을 확인할 수 있으며 패치 판별 정확도가 낮은 Galaxy S4 LTE, iPhone 6 C 모델의 경우가 동영상 판별 정확도도 낮았으며 Canon IXUS 160 모델도 정확도가 50%대로 낮은 것을 확인할 수 있다.

본 논문에서 제안하는 동영상 촬영 장치 판별 알고리즘의 경우 프레임 단위(또는 패치 단위)로 정확도를 판별한 후에 최대 빈도수를 갖는 것으로 촬영 장치를 판별하기 때문에 패치 판별의 정확도가 낮은 경우에도 올바르게 동영상 촬영 장치를 판별할 수 있는 것을 확인할 수 있으며 최종적으로 91.3%의 높은 정확도를 달성하였다.

표 5. 패치별 판별 정확도 및 동영상 판별 정확도
Table 5. Identification accuracy for each patch and videos

Device	Patch identification accuracy	Accurate video #	Video identification accuracy
Canon 650D	83%	10/10	100%
Canon EOS 500D	92%	10/10	100%
Canon EOS M	85%	10/10	100%
Canon IXUS 160	59%	5/10	50%
Nikon Coolpix S33	92%	10/10	100%
Nikon Coolpix S100	79%	3/4	75%
Panasonic DMC SZ1	76%	6/6	100%
SAMSUNG WB35F	76%	10/10	100%
GoPro	58%	10/12	83%
Lumix DWC LX100	97%	12/12	100%
G Pro2	82%	8/8	100%
G2	43%	7/7	100%
G3 Cat6	67%	8/10	80%
G4	65%	7/7	100%
Galaxy Grand Max	84%	7/7	100%
Galaxy Note2 A	82%	8/8	100%
Galaxy Note2 B	57%	7/7	100%
Galaxy Note3 A	53%	7/8	88%
Galaxy Note3 B	80%	7/7	100%
Galaxy Note3 C	62%	6/7	86%
Galaxy S4 LTE	40%	4/7	57%
Galaxy S5	51%	7/7	100%
iPhone 5S A	84%	8/8	100%
iPhone 5S B	77%	6/6	100%
iPhone 6 A	77%	7/7	100%
iPhone 6 B	59%	12/14	86%
iPhone 6 C	44%	4/7	57%
iPhone 6Plus	51%	5/7	71%
Vega secret note	63%	7/7	100%
Vu3	62%	8/8	100%
Galaxy alpha	82%	5/5	100%
Average	69.7%	231/253	91.3%

4.3 동일 브랜드의 동일 모델 판별 분석

기존의 특징 기반 촬영 장치에 대한 판별 연구들의 경우 브랜드나 모델에 대한 판별 보다는 개별 장치에 대한 판별이 가능하였다[12][13]. 그러나 딥러닝을 활용한 연구들의 경우는 개발 장치보다는 모델에 대한 판별이 가능하였고 개별 장치에 대한 판별은 상대적으로 정확도가 낮은 것으로 나타났다. Chen 등의 영상 촬영 장치 판별 결과를 분석해 보면 동일 모델에 대해서는 94.7% 정확도를 달성하였으나 동일 장치에 대한 판별에서 45.8%로 낮은 정확도를 보였다[9].

본 논문에서 적용한 고주파 통과 필터와 합성곱 신경망 모델도 패치별 식별 정확도는 69.7%로서 Chen 등의 결과보다는 높지만, 낮은 것을 확인할 수 있다. 또한, Galaxy Note3나 iPhone 6의 경우를 보면 동일 모델임에도 정확도 차이가 큰 것을 확인할 수 있다. 이는 딥러닝 기반의 모델들이 장치 식별보다는 모델 식별에 적합하도록 학습이 이루어지기 때문이고, 동일 장치들에 대한 구분이 어려워서 특정 장치로 학습이 편향되기 때문으로 판단된다.

그러나 본 논문에서 제안하는 최대 빈도수를 적용한 동영상 촬영 장치 판별 알고리즘을 통하여 장치에 대한 판별을 수행하는 경우 동일한 Galaxy Note2에 대해서는 100% 구분을 하였고, Galaxy Note3 모델에 대해서는 21개에서 2개를 제외하고 구분을 하였으며, 동일한 iPhone 5S 모델에 대해서는 100% 구분을 하였고, 동일한 iPhone 6 모델에 대해서는 28개 중 5개를 제외하고 구분을 하였다. 따라서 총 79개 동영상 중에서 72개를 올바르게 판별하여 동일 모델에 대해서도 91.1%의 장치 판별 정확도를 보였다.

기존 연구들이 영상을 대상으로 하고, 본 연구는 동영상을 대상으로 하여 차이는 있을 수 있겠지만 제안하는 알고리즘에서 도입된 최대 빈도수를 이용하여 촬영 장치를 판별하는 방법을 이용한다면, 기존의 연구들이 모델에 비해 낮은 성능을 보이는 장치 판별에 대해서도 충분히 정확도를 향상할 수 있을 것으로 판단된다.

4.4 다른 알고리즘과 성능 비교

본 논문에서 제안하는 알고리즘을 기존의 동영상 촬영 장치 판별 알고리즘과 비교를 수행하였고 그 결과를 표 6에 정리하였다.

촬영된 장치의 숫자가 다르고, 알고리즘과 분류 모델에 따라서 성능이 차이가 날 수밖에 없는 특징은 있으나, 제안한 모델이 더 많은 장치를 대상으로 하였음에도 불구하고 전통적인 판별 기법을 활용한 Milani 등[14] 및 Taspinar 등[15] 보다도 높은 정확도를 달성하였다. Hosler 등[16] 알고리즘 보다는 정확도가 낮은 것으로 보이지만 이는 사용한 동영상 촬영 장치의 수가 11개나 차이가 나기 때문으로 31개 장치를 사용할 경우 정확도가 급격히 떨어질 것으로 예상된다. Lee 등[13] 알고리즘과 비교할 때는 유사한 수의 장치를 활용하였으나 정확도가 상대적으로 낮은 것은, 영상 촬영 장치 및 동영상 촬영 장치 판별을 위한 딥러닝 기술들의 한계로 판단되며 성능 향상을 위해서는 추가적인 연구의 진행이 필요한 부분이다.

표 6. 다른 동영상 판별 장치 알고리즘과 성능 비교
Table 6. Performance comparison with other video capturing device identification algorithms

Algorithm	# of devices	Classification model	Accuracy
Milani et al.[14]	5	SVM	79.8%
Lee et al.[13]	30	Thresholding on probability density	96.0%
Taspinar et al.[15]	13		83.0%
Hosler et al.[16]	20	Deep learning	95.9%
Proposed	31		91.3%

Ⅴ. 결론 및 향후 과제

동영상 촬영 장치의 성능이 향상되고 가격이 저렴해짐에 따라서 누구나 손쉽게 동영상 콘텐츠를 생산하여 배포하고 있다. 그러나 불법적인 사용으로 인하여 사회적 문제점들이 야기되고 있으며 동영상 포렌식 기술에 대한 요구가 높다.

본 논문에서는 최근 관심을 받고 있는 딥러닝 기술을 적용하여 동영상 촬영 장치를 판별하기 위한 포렌식 알고리즘을 제안하였다. 영상 기반의 촬영

장치 판별 포렌식 알고리즘을 그대로 적용할 수 없기 때문에, 동영상 특성을 분석하여 96.2%의 정확도를 갖는 P 프레임 활용, 프레임 분할, 최대 빈도수 기반 판별을 하는 알고리즘을 설계하였고, 고주파 통과 필터와 합성곱 신경망을 기본으로 하는 딥러닝 모델을 계층별 성능을 분석하여 97.9%의 정확도를 갖는 4계층 모델을 동영상 촬영 장치 판별에 적합하도록 변형하여 적용하고 파라미터 최적화를 하였다.

특히, 기존 딥러닝에 기반하는 동영상 촬영 장치 판별 연구들에서는 실험하지 못한 총 31개의 방대한 촬영 장치 데이터를 활용하여 실험과 분석을 수행하였고, 91.3% 판별 정확도를 달성하였다. 또한, 제안하는 알고리즘은 모델이 아닌 장치에 대한 판별에 있어서도 91.1%의 정확도를 달성하여 기존 연구에 비하여 우수한 장치 판별 성능을 갖고 있음을 보였다.

본 연구의 실험에 있어서 동영상의 방대한 특성으로 인하여 학습에 걸리는 시간이 영상보다 훨씬 많이 요구된다. 따라서, 더 많은 장치들을 활용할 경우 처리 시간이 매우 오래 걸릴 수 있어서 동영상의 특정 영역만 추출하여 활용하는 형태의 접근 방법이 필요할 수 있다. 또한, 판별을 위하여 채택한 고주파 통과 필터와 합성곱 신경망 모델이 아닌 DenseNet 등과 같은 최신 기법을 도입하면 판별 성능이 더욱 향상될 수 있을 것으로 기대된다.

References

[1] Prosecution Service, "2019 analytical statistics on crime, crime occurrence and characteristics trend for 10 years", pp. 15, 2019.

[2] J. Lukas, J. Fridrich, and M. Goljan, "Digital camera identification from sensor pattern noise", IEEE Trans. on Information Forensics Security, Vol. 1, No. 2, pp. 205-214, Jul. 2006.

[3] M. Chen, J. Fridrich, and M. Goljan, "Source digital camcorder identification using ccd photo response non-uniformity", Proc. of SPIE Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents IX, San Jose, CA, pp. 1G-1H, Feb. 2007.

[4] S. Bayram, H. T. Sencar, N. Memon, and I. Avcibas, "Source camera identification based on CFA interpolation", IEEE International Conference on Image Processing 2005, Genova, Italy, pp. 69-72, Sep. 2005.

[5] K. Kuroki, K. Kurosawa, and N. Saitoh, "An approach to individual video camera identification", Journal of Forensic Sciences, Vol. 47, No. 1, pp. 97-102, Jan. 2002.

[6] A. Tuama, F. Comby, and M. Chaumont, "Camera model identification with the use of deep convolutional neural networks", 2016 IEEE International Workshop on Information Forensics and Security (WIFS), Abu Dhabi, United Arab Emirates, pp. 1-6, Dec. 2016.

[7] B. Wang, J. Yin, S. Tan, Y. Li, and M. Li, "Source camera model identification based on convolutional neural networks with local binary patterns coding", Signal Processing: Image Communication, Vol. 68, pp. 162-168, Oct. 2018.

[8] U. Kamal, A. M. Rafi, R. Hoque, S. Das, A. Abrar, and M. Hasan, "Application of DenseNet in camera model identification and post-processing detection", Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp. 19-28, May 2019.

[9] Y. Chen, Y. Huang, and X. Ding, "Camera model identification with residual neural network", 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, pp. 4337-4341, Sep. 2017.

[10] P. Yang, D. Baracchi, R. Ni, Y. Zhao, F. Argenti, and A. Piva, "A Survey of Deep Learning-Based Source Image Forensics", Journal of Imaging, Vol. 6, No. 9, pp. 1-24, Mar. 2020.

[11] S. Park and N. Kwak, "Analysis on the Dropout Effect in Convolutional Neural Networks", Proc. of Asian Conference on Computer Vision, Taipei, Taiwan, pp. 189-204, Nov. 2016.

[12] T. W. Oh, D. K. Hyun, K. B. Kim, and H. Y.

Lee, "Digital imaging source identification using sensor pattern noises", KIPS Trans. on Software and Data Engineering, Vol. 4, No. 12, pp. 561-570, Dec. 2015.

[13] S. H. Lee, D. H. Kim, T. W. Oh, K. B. Kim, and H. Y. Lee, "Digital video source identification using sensor pattern noise with morphology filtering", KIPS Trans. on Software and Data Engineering, Vol. 6, No. 1, pp. 15-22, Jan. 2017.

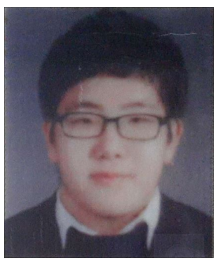
[14] S. Milani, L. Cuccovillo, M. Tagliasacchi, S. Tubaro, and P. Aichroth, "Video camera identification using audio-visual features", 2014 5th European Workshop on Visual Information Processing (EUVIP), Paris, France, pp. 1-6, Dec. 2014.

[15] B. Hosler, O. Mayer, B. Bayar, X. Zhao, C. Chen, J. A. Shackleford, and M. C. Stamm, "A Video Camera Model Identification System Using Deep Learning and Fusion", 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, United Kingdom, pp. 8271-8275, May 2019.

[16] S. Taspinar, M. Mohanty, and N. Memon, "Source camera attribution using stabilized video", 2016 IEEE International Workshop on Information Forensics and Security (WIFS), Abu Dhabi, United Arab Emirates, pp. 1-6, Dec. 2016.

저자소개

김 동 현 (Dong-Hyun Kim)



2016년 2월 : 금오공과대학교 컴퓨터소프트웨어공학과 (학사)
2016년 2월 ~ 현재 : 금오공과대학교 소프트웨어 공학과 석사과정
관심분야 : 이미지 처리, 포렌식, 딥 러닝

이 해 연 (Hae-Yeoun Lee)

1997년 : 성균관대학교 정보공학과 (학사)
1999년 : 한국과학기술원 전산학과 (공학석사)
2006년 : 한국과학기술원 전자전산학과 (공학박사)
2008년 ~ 현재 : 금오공과대학교 컴퓨터소프트웨어공학과 교수
관심분야 : Digital Forensics, Image Processing, IoT