딥러닝 기반 영상 조작 및 검출 기술 동향

□ 오병태 / 한국항공대학교

요약

다양한 목적으로 영상을 조작하려는 시도는 디지털 영상이 보편화되기 시작할 때부터 지속적으로 존재해 왔던 문제이며, 이러한 영상 조작의 유무를 검출하려는 시도 또한 지난 수십 년 동안 끊임없이 연구되어 왔다. 최근 빠르게 발전하는 인공지능 기술, 그 중에서도 딥러닝 기술을 이용하여 영상 조작을 검출하는 기술이 다양하게 발전되고 있지만, 한편으로는 딥러닝 기술을 이용하여 조작을 보다 정교하게 진행하거나 검출을 회피하려는 기술 또한 빠르게 발전하고 있다. 본 고에서는 영상을 조작하고, 검출하고 회피하는 기술 동향에 대하여 종합적으로 소개하고, 특히 딥러닝 기반의 기술이 각각의 영역에서 어떻게 적용되고 발전하고 있는지에 대하여 면밀히 살펴보고자 한다.

Ⅰ. 서 론

카메라로 우리의 일상을 낱낱이 기록하고, 기록된 영상을 소셜 네트워크 서비스(SNS) 등에 공유하는 것은 점차 우리 생활의 일부가 되어가고 있다. 이러한 디지털 기반의 삶의 변화는 점차 가속화될 것으로 전망되고 있다. 하지만, 이러한 기술적 편리함을 악용하여 오히려 우리 삶을 불편하게 하는 좋지 않은 사례가 최근 급증하고 있다. 예를 들면 누군가가 인터넷에 공유되어 있는 사진을 악의적인 목적으로 조작하여 퍼트리거나, 혹은 가짜뉴스의 생성 및 확산을 위해 원 영상을 가짜뉴스에 맞게 조작하는 사례 등이다. 이러한 문제는 단순히 한 개인의 명예훼손에 그치는 것이 아니라, 기업 혹은 단체 등에 대한 조작된 정보의 유포를 통해 정치적 혹은 경제적 측면에서 심각한 악영향을 끼치기도 한다. 특히, 최근 빠르게 발전하고 있는 영상 편집 소프트웨어의 확산에 따라 영상 조작에 대한 접근이 더욱더 용이해지면서 이러한 사회문제는 점차 심각해질 것으로 예상된다. 따라서, 이러한 불필요한 사회적 비용을 줄이기 위하여 조작되어 유포되는 영상을 즉각적으로 검출하여 인지시키는 기술 개발이 필요하다.

디지털 영상에 대한 조작 여부 검출은 영상 포렌식

(Image forensic)이란 이름으로 지난 2000년 이전부터 활발하게 연구되어 왔으며[1, 2], 최근 딥러닝 기술의 폭발적인 성능향상에 발맞추어 영상 포렌식 분야에서도 딥러닝 기술을 기반으로 한 검출 기술이 활발하게 연구되고 있다[3, 4]. 하지만, 포렌식 분야의 특성상 딥러닝 등의 기술발전은 역설적으로 영상 조작의 정밀도를 높여주는 동시에 조작 검출에 대한 회피를 목적으로 개발되는 안티포렌식(Anti-forensic) 기술의 성능을 향상시키는 부작용을 가지고 있어, 현재에도 끊임없이 영상 조작을 하려는 부류와 이를 검출하려는 부류 간의 기술적 다툼이 일어나고 있다.

본 고에서는 조작을 검출하려는 포렌식 기술과 이를 회피하려는 안티포렌식 기술 동향에 대하여 각각 상세히 살펴보고, 향후 영상 포렌식 분야가 나가가야 할 방향에 대해서 논의해 보고자 한다.

Ⅱ. 영상 조작 기술과 검출 기술의 발전

영상을 조작하는 방식은 그 목적과 방법에 따라 다양하게 분류할 수 있다. 얼굴이나 표정, 자동차 번호판 등, 영상 내 원하는 콘텐츠를 임의로 변경하여 조작을 진행하는 경우도 있고, 영상 내 다른 영역이나 다른 영상의 일부를 복사하여 붙이는 복사-붙여넣기(copy-and-paste) 방식으로 조작을 진행할 수도 있으며, 동영상의 경우에는 일부 프레임을 삭제하거나 순서를 바꾸는 방식 등으로 원 콘텐츠를 조작할 수 있다. 아래 〈그림 1〉에서는 일반적으로 많이 사용되는 영상 조작의 방법들을 보여주고 있다. 사실 이러한 모든 방식들은 영상 편집의 일종으로서 최근 빠르게 발전하고 있는 포토샵, 프리미어 등의 영상 편집 툴을 이용하면 더욱 더 정교하게 영상 조작이 가능하다.

이렇게 조작된 영상을 검출하기 위해 그동안 다양한 기술들이 소개되어 왔다. 본 고에서는 딥러닝 기술이 적용되기 전의 전통적인 조작 검출 기술과 최신 기술로서 딥러닝을 활용한 조작 검출 방식에 대하여 살펴본다.

1. 전통적인 영상 조작 검출 기술

전통적인 방법을 적용하여 영상 조작을 검출하는 기

<그림 1> 다양한 영상 조작 방식의 예

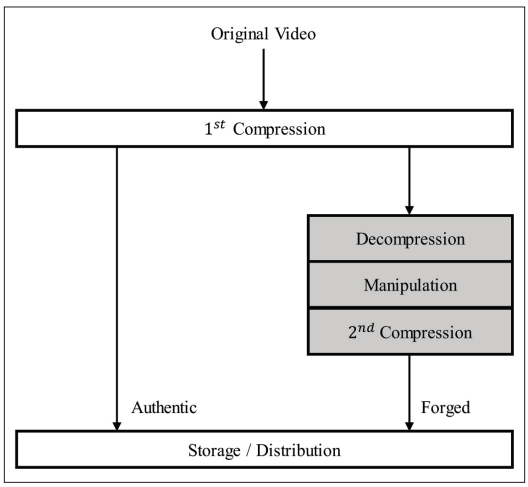
술은 주로 임의의 조작 과정에서 생기는 특이한 패턴이나 새롭게 발생하는 규칙의 불규칙화, 혹은 불규칙의 규칙화를 검출하는 방식을 주로 사용하고 있다.

먼저, 초창기부터 많이 사용되던 방식으로서 카메라의 고유 잡음 및 패턴을 이용하는 방식이 있다. 카메라에서 촬영된 영상에서는 카메라의 하드웨어적 특성에 의하여 영상 정보를 기록하는 과정에서 카메라 고유의 잡음 및 패턴을 내포하게 된다. 디지털 카메라의 경우 수많은 포토다이오드를 이용하여 빛의 정보를 센싱하게 되는데, 하드웨어 제조과정에서의 특징 및 한계에 의하여 모든 센서가 동일하도록 만드는 것은 불가능하다. 따라서, 영상 포렌식 초창기에는 이와 같은 과정에서 발생하는 카메라 잡음을 photo response nonuniformity (PRNU)라고 하고, 이를 영상의 조작 여부 검출에 적극적으로 이용하였다. 또한 PRNU 정보는 각 카메라의 제조과정에서 발생하기 때문에 카메라의 identity를 찾는데도 사용되기도 한다[5].

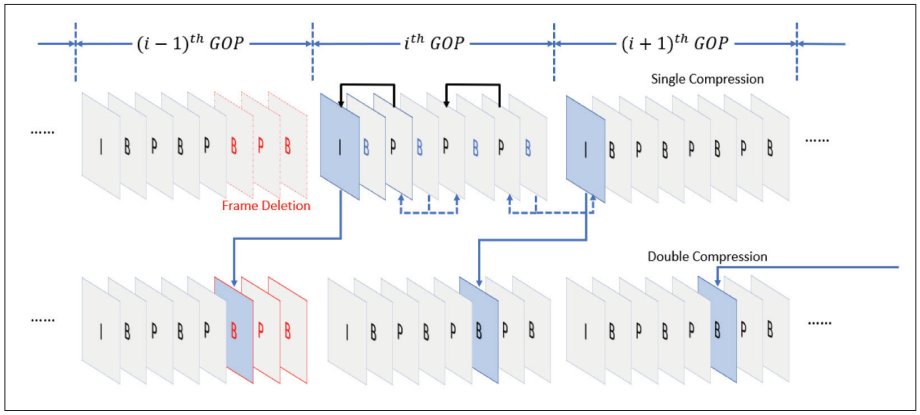
PRNU 잡음 이외에 카메라에서 컬러 영상을 센싱하기 위해 사용하는 color filter array(CFA) 패턴을 가지고 영상의 조작 여부 및 카메라 identity를 찾는데 활용하기도 한다. 예를 들면 다수의 카메라가 컬러영상 센싱 시 보편적으로 사용하는 Bayer 패턴을 동일하게 적용했더라도, 디모자이킹(demosaicing) 방식을 적용하는 과정에서 각 카메라 제조사만의 기술이 적용되는데, 여기서 발생하는 미세한 차이를 바탕으로 분류하는 방식이 주로 사용되고 있다[6].

영상 조작시 생기는 기하학적인 불일치나 광원, 그림자 등의 물리적인 현상을 분석하여 영상 내의 불일치를 찾는 방식도 활발하게 연구되었다. 대표적인 영상 조작 방식 중 하나인 복사-붙여넣기 방식 등을 이용하는 경우 이러한 기하학적 불일치 등이 빈번하게 일어나게 된다. 특히 3차원 영상의 2차원 사영 과정에서 생기는 고유의 형태 변화에 대하여, 영상 내 정보를 그대로 이용하는 복사-붙여넣기 조작 등은 이를 완전히 고려하기 어렵기 때문에 조작 흔적을 남기게 된다. 하지만, 이러한 영상 내 불일치를 찾는 과정이 매우 복잡하고 부정확하여 일반적인 경우 높은 검출 정확도를 기대하기는 어렵다.

또다른 전통적인 방식으로 크게 각광받는 방식이 이중 압축 혹은 다중 압축 여부를 판별하는 방식이다. 일반적으로 카메라로 영상을 취득하는 경우 이를 원 데이터 그대로 사용하는 경우는 극히 드물고, 대부분의 영상 획득 장치에서는 영상 획득과 동시에 바로 압축을 진행하게 된다. 따라서, 일반적으로 유통되는 대부분의 영상은 압축된 영상 혹은 동영상이라고 생각할 수 있다. 하지만, 특정한 압축 기술로 압축되어 있는 영상을 조작하기 위해서는 〈그림 2〉와 같이 압축이 된 영상을 압축해제하고, 원하는 조작을 진행한 후, 다시 재압축하는 이중 압축 과정을 반드시 거쳐야만 한다. 따라서, 이중압축 여부를 판단하는 방식이 영상의 진위여부를 판단하는 하나의 가장 중요한 기술 분야가 되었다[1, 2]. JPEG 등의 영상 압축의 경우, 압축과정에서 발생하는 라운딩 에러 혹은 절단(truncation) 에러를 살펴보기도 하고, 이중 압축시 적용하는 양자화 계수의 불일치에 따른



<그림 2> 영상 조작시 발생하는 이중 압축 과정



<그림 3> 영상 삭제시 발생하는 GOP 구조의 변화

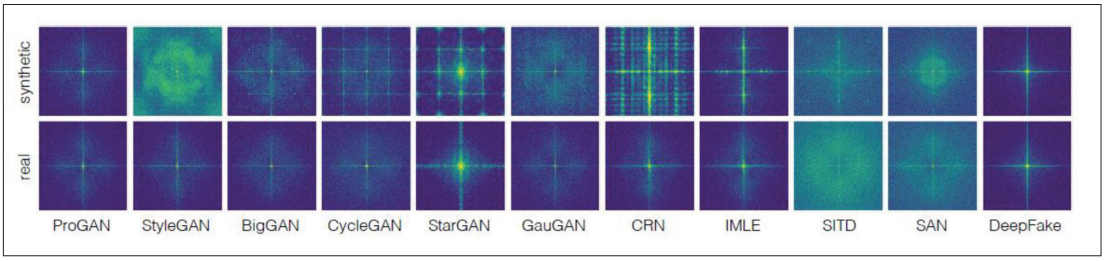
히스토그램 변화 등을 이용하는 방식도 제안되었다. 동영상의 경우에는 여러 조작에 의하여 〈그림 3〉과 같이 Group of Picture(GOP)의 특성이 바뀌는 것을 파악하는 방식, 각종 압축 신택스의 변화를 측정하는 방식 등 다양한 방식이 소개되었다[7-10].

2. 딥러닝 기반의 영상 조작 검출

2000년대 이후 학습 기반의 방식이 영상처리 및 비전 분야에서 점차 많은 관심을 받게 되었고, 2010년대 이후에는 딥러닝 기술을 기반으로 한 다양한 기술들이 폭넓은 연구 분야에서 중요하게 사용되고 있다. 영상 조작 검출 분야에서도 이러한 연구 추세에 맞추어 최근 딥러닝 기반 방식들이 매우 활발하게 연구되고 있다.

먼저 영상의 이중 압축 여부를 판단하는 방법으로 딥러닝 기술을 적용하는 기술들을 생각해 볼 수 있다. 기본적으로 앞서 설명한 전통적인 방식들에서 사용했던 특징들을 딥러닝 방식을 이용하여 보다 정밀하게 분류하는 방식들이 주로 제안되었다. 즉 라운딩 및 절단 에러를 이용하거나, 이를 바탕으로 통계적인 특징을 추출하는 방식, 다양한 양자화 환경을 고려한 방식 등이 그 예이다[11-13]. 특히 이중압축 여부에 대한 판단 과정에서 첫 번째 압축과 동일한 환경 및 파라미터를 가지고 두 번째 압축을 진행하는 경우 딥러닝 이전의 전통적인 방법으로는 이를 판단하는 것이 큰 어려움이 있었다. 하지만 기계학습 및 딥러닝 기술을 바탕으로 이를 보다 효율적으로 판단하는 기술이 소개되기도 하였다 [14, 15].

영상 내 이상 패턴이나 잡음을 감지하는 방식은 보다 직관적으로 딥러닝 기반 기술을 적용할 수 있는 접근방식이다. 충분한 양의 원영상 및 조작영상 DB를 통하여 기존 많이 사용되고 있는 ResNet, InceptionNet 등의 높은 분류 기술을 사용하여 조작 영상을 검출하는 방식의 우수성이 입증되었다. 하지만, 이러한 방식들은 영상 내 픽셀 도메인의 정보만을 이용하는 것에 그치게 될 수 있어, 적대적 생성 신경망(Generative adversarial network, GAN)과 같이 영상 도메인에서 매우 정교하게 영상을 조작하는 경우 기존 전통적인 검출 방식은 한계가 명확하였다. 이러한 점을 극복하기 위하여 영상의 주파수 분석을 통하여 검출하는 방식들도 활발하게 제안되었다. 대표적으로 아래 그림과 같이 웨이블릿 변환 혹은 푸리에 변환, 그 밖에 다양한 변화 기술들을 딥러닝 구조를 사용하여 구현하여 이를 바탕으로 조작검출



<그림 4> 주파수 분석을 통한 GAN 적용 영상 검출 사례 [19]

을 하는 다양한 방식들이 소개되었다. 이러한 기술들은 영상을 조작하는 입장에서 픽셀 도메인 이외의 변환 도메인까지 신경을 써서 조작을 하기 어렵다는 점 때문에 상당히 우수한 성능을 보여주고 있다[16-19].

조작영상을 검출하기 위한 학습 기반의 기술개발을 위해서는 반드시 우수한 데이터가 확보되어야 한다. 영상 조작 및 검출 분야에서도 최근 좋은 학습 DB를 구축하려는 노력들이 많이 있다. 기존 수많은 방법으로 원영상에 대하여 전통적인 방식으로 조작을 진행하여 DB를 구축하였으나[20, 21], 최근에는 GAN과 같은 방식으로 보다 정교하면서도 회피하기 힘들도록 만든 DB가 널리 쓰이고 있다[22, 23].

Ⅲ. 조작검출 회피를 위한 안티포렌식 기술

안티포렌식 기술은 영상 조작 검출을 위해 개발된 포렌식 알고리즘을 회피하기 위해 만들어진 기술로서, 다양한 목적으로 포렌식 기술의 발전과 함께 지속적으로 발전하고 있는 분야이다. 따라서 특정한 기술을 이용하여 포렌식 기술이 개발되면 이를 회피하게 되는 안티포렌식 기술이 뒤따라 개발되기 때문에, 포렌식 기술 개발을 진행할 때는 적용가능한 안티포렌식 기술을 모두 무력화 시키거나 적용하기 힘든 방식으로 개발되어야 하는 어려움이 따르게 된다. 따라서, 포렌식 기술과 안티포렌식 기술을 오랫동안 상호 대응하면서 지속적으로 발전하고 있다. 본 장에서는 이러한 안티포렌식 기술을 앞서와 같이 전통적인 기법과 딥러닝 기반의 학습 기반 방식으로 나누어서 설명한다.

1. 전통적인 안티포렌식 기술

일반적으로 포렌식 기술에서는 영상 조작시 발생하는 새로운 패턴 혹은 패턴의 변화를 감지하여 검출하는 것이 일반적이다. 따라서 안티포렌식 기술에서는 이러한 패턴 변화를 숨기거나 보정하는 방식으로 많이 기술 개발되었다.

가장 단순하면서도 효과적인 안티포렌식 기술은 일반적으로 영상처리 등에 사용되는 필터링 기법을 이용하는 방식이다. 그 중에서 중앙값 필터(Median filter)는 간단한 알고리즘의 동작방식에 비하여 패턴을 감추는데 매우 우수한 성능을 가지기 때문에 많이 사용되어 온 기술이다. 특히, 중앙값 필터는 이중압축 여부를 숨기기 위하여 압축과정에서 생기는 블록 잡음 등을 없애는데 효과적이었다. 하지만, 단순하게 필터를 적용하면 특유의 streaking 잡음 등이 생겨날 수 있기 때문에, 이를 보정할 수 있는 여러 필터들이 지속적으로 소개되

었다[24,25].

이밖에도 영상 내 대비(contrast)를 조정하는 방식, 디더링(dithering) 등을 이용하여 영상 내 복사-붙여넣기 하는 부분을 매끄럽게 하는 방식, 히스토그램 리매핑(remapping), 영상 크로핑(cropping) 및 리샘플링(resampling) 등 기존 영상처리 기반의 여러 방식들이 다양하게 소개되고 연구되었다[26].

2. 딥러닝 기반 안티포렌식 기술

최근 딥러닝의 발전과 더불어 딥러닝 기반의 안티포렌식 기술이 본격적으로 소개되었으며, 기존 간단한 필터 방식으로 안티포렌식을 진행했던 것과는 달리 훨씬 더 정교한 조작이 가능하게 되었다. 특히 적대적 생성 신경망인 GAN을 이용하여 생성 모델을 안티포렌식 기술로 사용하는 방식이 널리 사용되고 있다.

예를 들면 앞서 언급한 중앙값 필터의 경우, 정교하게 설계된 검출기에 의해 중앙값 필터 적용 유무가 검출될 수 있다. 따라서, 중앙값 필터를 적용했다는 사실을 감추기 위해 GAN을 활용하여 이러한 흔적을 감추는 기법이 소개되기도 하였다[27]. 이중 압축 기반의 포렌식 방식에 대응하기 위하여 압축의 흔적을 제거하는 기법도 널리 사용되는 방식이다. 즉, 한 번 압축된 영상에 대하여 조작을 진행한 후, GAN을 이용하여 영상 내 내포되어 있는 압축 흔적을 지움으로써, 이후 두 번째 압축을 적용했을 때, 이 영상이 한 번 압축된 영상과 큰 차이가 없도록 보이게 만드는 방식도 소개되었다[28]. 최근에는 다양한 종류의 딥러닝 기반 포렌식 검출기에 대응하기 위해, 이러한 검출기들을 동시에 분석하여 GAN을 이용하여 이들을 모두 회피할 수 있는 학습 시스템을 구축한 연구도 소개되었다[29].

Ⅳ. 결 론

지금까지 영상의 조작 기술 및 조작을 검출하는 포렌식 기술, 그리고 포렌식 기술을 회피하기 위한 안티포렌식 기술에 대하여 살펴보고, 특히 최근 크게 각광받고 있는 딥러닝 기반의 포렌식 및 안티포렌식 기술에 대하여 살펴보았다. 점차 정밀해지는 조작 및 회피 기술, 그리고 영상 콘텐츠의 파급효과가 매우 큰 이 시점에서 보다 강인하면서도 회피하기 힘든 포렌식 기술에 대한 수요가 점차 커질 것으로 예상된다. 기존 단순한 방식으로 조작되어 온 방식에서 벗어나서 매우 정교한 방식으로 설계된 조작 영상, 그리고 포렌식 기술이 적용될 것을 예상하고 설계된 안티포렌식 기술에 대응하기 위하여, 향후 연구에서는 이러한 회피기술을 무력화시킬 수 있는 전혀 새로운 방식의 카운터 안티포렌식 기술에 대한 연구개발이 반드시 필요하다.

참 고 문 헌

[1] Piva, Alessandro. "An overview on image forensics." International Scholarly Research Notices (2013).

[2] Milani, Simone, et al. "An overview on video forensics." APSIPA Transactions on Signal and Information Processing (2012).

[3] Yang, Pengpeng, et al. "A survey of deep learning-based source image forensics." Journal of Imaging 6.3 (2020): 9.

[4] Castillo Camacho, Ivan, and Kai Wang. "A Comprehensive review of deep-Learning-based methods for image forensics." Journal of Imaging 7.4 (2021): 69.

[5] Filler, Tomás, Jessica Fridrich, and Miroslav Goljan. "Using sensor pattern noise for camera model identification." 2008 15th IEEE International Conference on Image Processing. IEEE, 2008.

[6] Ferrara, Pasquale, et al. "Image forgery localization via fine-grained analysis of CFA artifacts." IEEE Transactions on Information Forensics and Security 7.5 (2012): 1566-1577.

[7] Li, Bin, Yun Q. Shi, and Jiwu Huang. "Detecting doubly compressed JPEG images by using mode based first digit features." 2008 IEEE 10th Workshop on Multimedia Signal Processing. IEEE, 2008.

[8] Li, Jixian, et al. "Double JPEG compression detection based on block statistics." Multimedia Tools and Applications 77.24 (2018): 31895-31910.

[9] Jiang, Xinghao, et al. "Detection of double compression in MPEG-4 videos based on Markov statistics." IEEE Signal processing letters 20.5 (2013): 447-450.

[10] Huang, Fangjun, Jiwu Huang, and Yun Qing Shi. "Detecting double JPEG compression with the same quantization matrix." IEEE Transactions on Information Forensics and Security 5.4 (2010): 848-856.

[11] Huang, Xiaosa, Shilin Wang, and Gongshen Liu. "Detecting double JPEG compression with same quantization matrix based on dense CNN feature." 2018 25th IEEE International Conference on Image Processing (ICIP). IEEE, 2018.

[12] Bakas, Jamimamul, Anil Kumar Bashaboina, and Ruchira Naskar. "MPEG double compression based intra-frame video forgery detection using CNN." 2018 International Conference on Information Technology (ICIT). IEEE, 2018.

[13] Hong, Jin Hyung, Yoonmo Yang, and Byung Tae Oh. "Detection of frame deletion in HEVC-Coded video in the compressed domain." Digital Investigation 30 (2019): 23-31.

[14] Jiang, Xinghao, et al. "Detection of HEVC double compression with the same coding parameters based on analysis of intra coding quality degradation process." IEEE Transactions on Information Forensics and Security 15 (2019): 250-263.

[15] Uddin, Kutub, Yoonmo Yang, and Byung Tae Oh. "Double compression detection in HEVC-coded video with the same coding parameters using picture partitioning information." Signal Processing: Image Communication (2022): 116638.

[16] Zhang, Xu, Svebor Karaman, and Shih-Fu Chang. "Detecting and simulating artifacts in GAN fake images." 2019 IEEE International Workshop on Information Forensics and Security (WIFS). IEEE, 2019.

[17] Qian, Yuyang, et al. "Thinking in frequency: Face forgery detection by mining frequency-aware clues." European Conference on Computer Vision. Springer, Cham, 2020.

[18] Frank, Joel, et al. "Leveraging frequency analysis for deep fake image recognition." International Conference on Machine Learning. PMLR, 2020.

[19] Wang, Sheng-Yu, et al. "CNN-generated images are surprisingly easy to spot... for now." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.

[20] Zampoglou, Markos, Symeon Papadopoulos, and Yiannis Kompatsiaris. "Detecting image splicing in the wild (web)." 2015 IEEE International Conference on Multimedia & Expo Workshops. IEEE, 2015.

[21] Korus, Paweł, and Jiwu Huang. "Multi-scale analysis strategies in PRNU-based tampering localization." IEEE Transactions on Information Forensics and Security 12.4 (2016): 809-824.

[22] Guan, Haiying, et al. "MFC datasets: Large-scale benchmark datasets for media forensic challenge evaluation." 2019 IEEE Winter Applications of Computer Vision Workshops. IEEE, 2019.

[23] Rossler, Andreas, et al. "Faceforensics++: Learning to detect manipulated facial images." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019.
[24] Kang, Xiangui, et al. "Robust median filtering forensics using an autoregressive model." IEEE Transactions on Information Forensics and Security 8.9 (2013): 1456-1468.
[25] Kirchner, Matthias, and Rainer Bohme. "Hiding traces of resampling in digital images." IEEE Transactions on Information Forensics and Security 3.4 (2008): 582-592.
[26] Qureshi, Muhammad Ali, and El-Sayed M. El-Alfy. "Bibliography of digital image anti-forensics and anti-anti-forensics techniques." IET Image Processing 13.11 (2019): 1811-1823.
[27] Kim, Dongkyu, et al. "Median filtered image restoration and anti-forensics using adversarial networks." IEEE Signal Processing Letters 25.2 (2017): 278-282.
[28] Luo, Yingmin, et al. "Anti-forensics of JPEG compression using generative adversarial networks." 2018 26th European Signal Processing Conference (EUSIPCO). IEEE, 2018.
[29] Zhao, Xinwei, Chen Chen, and Matthew C. Stamm. "A Transferable anti-forensic attack on forensic CNNs using a generative adversarial network." arXiv preprint arXiv:2101.09568 (2021).

필 자 소 개

오 병 태

- 2003년 : 연세대학교 전기전자공학부 학사
- 2009년 : Univ. of Southern California(USC), Dept. of Electrical Eng. 석사 및 박사
- 2009년 ~ 2013년 : 삼성종합기술원 전문연구원
- 2013년 ~ 현재 : 한국항공대학교 항공전자정보공학부 교수
- 주관심분야 : 영상처리, 비디오압축, 영상 포렌식