딥러닝 기반 얼굴 위변조 검출 기술 동향

□ 김원준 / 건국대학교

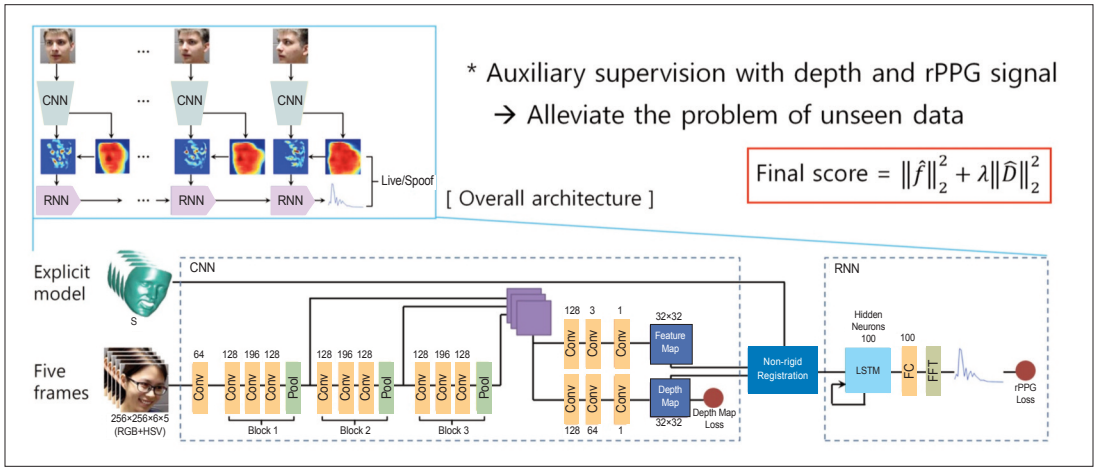
요약

최근 생체 정보를 이용한 사용자 인증 기술이 발전하면서 이를 모바일 기기에 적용하는 사례가 크게 증가하고 있다. 특히, 얼굴 기반 인증 방식은 비접촉식이며 사용이 편리하여 적용 범위가 점점 확대되고 있는 추세이다. 그러나, 사용자의 얼굴 사진이나 동영상 등을 이용한 위변조가 용이하기 때문에 모바일 기기 내 보안 유지에 어려움을 야기한다. 본 고에서는 이러한 문제를 해결하기 위해 최근 활발히 연구되고 있는 심층신경망 기반 얼굴 위변조 검출 연구의 최신 동향을 소개하고자 한다. 먼저, 기본 합성곱 신경망 구조부터 생성 모델 기반의 위변조 검출 방법까지 다양한 신경망 구조를 이용한 위변조 검출 방법에 대해 설명한다. 또한, 심층신경망 학습을 위해 사용되는 얼굴 위변조 데이터셋에 대해서도 간략히 살펴보고자 한다.

I. 서 론

최근 스마트폰 등 다양한 모바일 기기에 생체 정보를 기반으로 한 인증 시스템이 적용되면서 많은 사용자들이 기존 패스워드 입력 방식 대신 얼굴이나 지문 등을 이용하여 본인 인증을 수행하고 있다. 또한, 은행 거래 및 모바일 결제 응용에도 생체 인증이 적용되면서 고성능 보안 유지가 가능한 인증 시스템 개발에 대한 관심이 크게 증가하고 있다. 사용자 생체 정보 중 얼굴 영상의 경우 등록 및 인증 과정이 비접촉식으로 이루어지며 그 과정이 간결하기 때문에 가장 널리 사용되고 있는 생체 정보이다. 그러나 사회 관계망 등을 이용한 인위적 획득이 매우 쉽고, 사진 출력 및 동영상 제작을 이용한 위변조에 취약한 문제점이 있다. 이를 해결하기 위해 컴퓨터 비전 분야에서는 별도의 센서 등을 이용한 하드웨어적 위변조 검출이 아닌 소프트웨어적 솔루션에 대해 꾸준히 연구해오고 있다.

얼굴 위변조 검출 방법은 영상 내 텍스처(Texture) 정보를 이용한 검출 방법과 심층학습을 이용한 검



<그림 1> 얼굴 위변조 검출을 위해 보조 정보를 함께 학습하는 심층신경망 구조[3]

출 방법으로 크게 나눌 수 있다. 전자의 경우, 모든 픽셀 위치에서 주변 픽셀 값과의 비교 결과를 이진 패턴으로 인코딩하는 기술자(Local Binary Pattern, LBP)[1]를 이용하여 실제 얼굴과 위변조 얼굴의 차이를 구별하고자 노력하였다. 그러나 고해상도 영상 제작이 가능해지면서 실제 얼굴과 위변조 얼굴 간 미세한 표면 차이를 픽셀 밝기 값의 관계만으로 구별하기 매우 어려우며, 다양한 위변조 형태를 효과적으로 학습하는데 한계가 있다. 한편, 영상 인식 분야에서 뛰어난 성과를 보여준 심층신경망(Deep Neural Network)을 얼굴 위변조 검출에 적용하고자 하는 연구가 몇몇 연구자들에 의해 시작되고 있다. 얼굴 위변조는 주어진 입력 영상이 실제 얼굴인지 아닌지를 구별하는 이진 분류(Binary Classification) 문제로 간주할 수 있기 때문에 영상 인식에 널리 사용되는 합성곱 신경망 구조를 기본적으로 적용할 수 있다. 즉, 입력 영상은 적층 구조의 합성곱 계층을 반복적으로 통과하면서 얼굴 위변조 검출을 위한 잠재 특징(Latent Feature)으로 압축되며 이를 기반으로 심층신경망 기반 방법은 기존 영상 특징 대비 검출 정확도를 효과적으로 개선하였다. 이러한 심층신경망 기반 방법은 다양한 위변조 공격 형태를 모델 용량 증가 없이 학습할 수 있는 장점이 있으며, 깊은 적층 구조(Deeply Stacked Architecture)를 기반으로 위변조 얼굴 영상의 다양한 비선형적 변형 또한 성공적으로 학습할 수 있다. 그러나 학습 데이터셋에 존재하지 않는 위변조 공격에 대해서는 매우 취약하며, 기존 영상 텍스처 기반 위변조 검출 방법 대비 수행 속도가 느린 문제점이 있다. 또한, 데이터셋 구축에 많은 인력과 시간이 소요되기 때문에 큰 비용이 필요하며, 새로운 위변조 공격 형태가 추가될 때마다 학습 과정이 다시 수행되어야 하는 한계가 있다. 그럼에도 불구하고 뛰어난 정확도 및 확장성으로 인해 많은 연구자들이 다양한 심층신경망 구조를 기반으로 한 얼굴 위변조 검출 연구를 진행하고 있다.

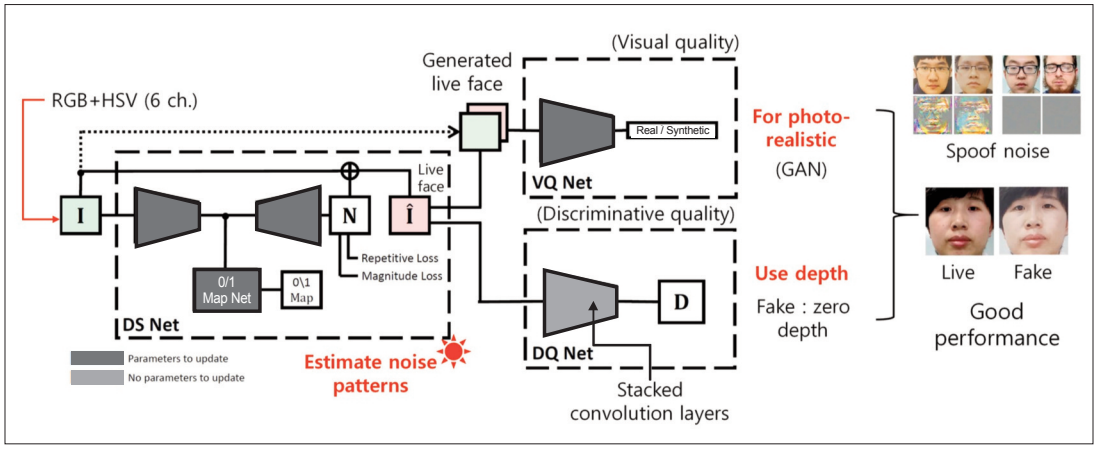
본 고에서는 이와 같이 심층신경망을 기반으로 한 얼굴 영상 기반 위변조 검출의 최신 기술 동향을 살펴보고자 한다. II장에서는 최근 주요 학회에서 발표된 논문을 중심으로 심층신경망 기반 얼굴 위

변조 검출 기술의 연구 방향에 대해 자세히 설명한다. III장에서는 얼굴 위변조 검출 기술 개발을 위한 데이터셋에 대해 간략히 살펴보고, 마지막으로 IV장에서 결론을 맺는다.

II. 심층신경망 기반 얼굴 위변조 검출 기술 동향

최근 영상 인식 분야에서 뛰어난 성능 향상을 입증한 심층학습 기술을 얼굴 위변조 검출에 적용하려는 시도가 늘고 있다. 기본적으로 입력 영상을 신경망에 입력으로 하여 실제 얼굴인지 아닌지에 대한 점수를 출력하는 구조를 기반으로 적용이 시작되었으며, 생성 모델(Generative Model)을 기반으로 위변조 잡음 정도를 측정하는 방법도 소개되고 있다. 가장 최근에는 학습 데이터셋에 포함되지 않은 새로운 위변조 영상을 검출하기 위해 보조 정보(Auxiliary Information)를 추가로 학습할 수 있는 신경망 구조, 다중 채널 영상(예를 들어, RGB 칼라, 근적외선(Near Infrared, NIR), 깊이(Depth) 영상 등)을 입력으로 학습할 수 있는 다중 브렌치(Multi-branch) 구조의 신경망 등 고성능 얼굴 위변조 검출을 위한 다양한 심층신경망 구조가 개발되고 있다.

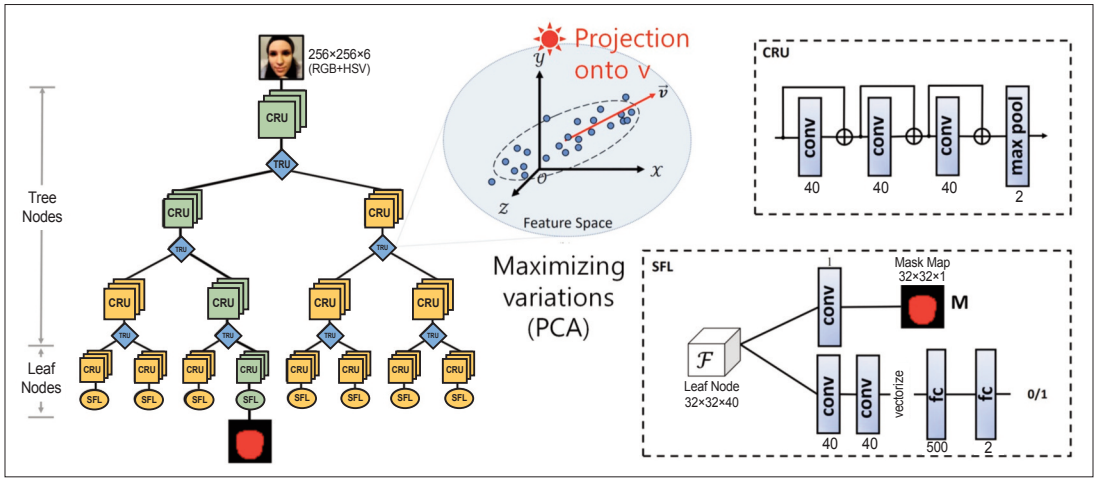
가장 먼저 영상 인식 분야에서 널리 사용되고 있는 합성곱 신경망(Convolution Neural Network, CNN) 구조가 얼굴 위변조 검출에 적용되기 시작했다. 자세히 살펴보면, Atoum[1] 등은 먼저, 입력 얼굴 영상을 지역적 영역으로 분할하고 각 영역에 대해 합성곱 신경망을 적용하여 위변조 점수를 측정한다. 또한, 3D 얼굴 복원 알고리즘을 통해 추출된 결과 영상을 기반으로 생성 모델 신경망을 학습하고, 도출한 잠재 특징으로 위변조 점수를 계산한다. 이와 같이 계산된 두 점수에 대한 평균값을 이용하여 최종 위변조 여부를 판별하는 알고리즘으로 기존 영상 텍스처 기반 방법 대비 큰 성능 향상을 달성하였다. 이와 더불어 정지 영상이 아닌 얼굴 동영상을 효과적으로 학습할 수 있도록 3차원 합성곱 연산을 적용한 방법도 제안되었다[2]. Liu[3] 등도



<그림 2> 얼굴 위변조 잡음 생성을 기반으로 한 위변조 여부 검출 신경망 구조[5]

얼굴 동영상을 이용하여 위변조 여부를 판별하기 위해 3차원 합성곱 연산 대신 2차원 합성곱 연산을 통해 추출된 잠재 특징을 장단기 기억(Long Short-Term Memory, LSTM) 모듈을 이용하여 시계열 관계를 효과적으로 추정하여 위변조 여부 검출에 사용하는 방법을 제안하였다. 특히, 얼굴 위변조 검출 문제를 단순 이진 분류 문제가 아닌 회귀(Regression) 관점에서의 위변조 정도(Degree)를 측정하는 문제로 정의하여 보조 정보(즉, 얼굴 깊이 영상과 원격광혈류측정(Remote Photoplethysmography, rPPG) 신호)를 함께 학습할 수 있는 신경망 구조를 제안하였다(〈그림 1〉 참조). [3]의 방법을 시작으로 깊이 영상과 원격광혈류측정 신호는 최근 연구에서 폭넓게 적용되고 있다. Yang[4] 등 또한 전체 얼굴 동영상에 대하여 장단기 기억 모듈을 기반으로 전역적 특징을 학습하고, 관심 영역을 기반으로 추출된 지역적 얼굴 영역에서 특징을 함께 이용하여 얼굴 위변조 검출을 수행한다. 한편으로, 위변조 얼굴 영상은 실제 얼굴 영상과 위변조 잡음으로 구성되어 있다는 가정하에 생성 모델을 이용하여 위변조 잡음(Spoof Noise) 이미지를 생성하는 연구도 진행되었다[5]. 추정된 잡음 이미지를 제거하여 생성한 실제 얼굴 영상의 자연스러움을 측정할 수 있는 적대적 손실(GAN Loss) 함수와 [3]과 마찬가지로 얼굴 깊이 영상을 기반으로 생성된 실제 얼굴 영상의 분별력을 측정할 수 있는 손실 함수를 기반으로 고성능 위변조 여부 판별이 가능하다. 〈그림 2〉는 위변조 잡음 기반 얼굴 위변조 검출 심층신경망 구조를 보여주고 있으며, 실제 얼굴에 대해 추정한 위변조 잡음 이미지에는 검출 신호가 거의 나타나지 않음을 확인 할 수 있으며(오른쪽 위 참조), 이를 통해 위변조 여부를 효과적으로 판별할 수 있다.

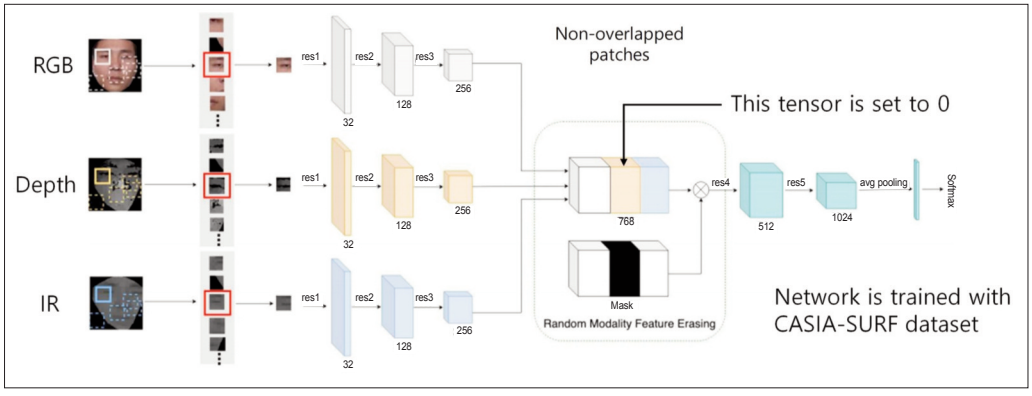
심층신경망 기반 얼굴 위변조 검출 방법들은 학습된 데이터셋과 동일한 공격 유형에는 매우 강인한 동작이 가능하지만, 학습 데이터셋에 존재하지 않는 위변조 영상에 대해 검출 성능이 저하되는 단점이 있다. 이러한 문제를 개선하기 위한 방법으로



<그림 3> 나무 구조와 심층신경망을 이용한 얼굴 위변조 검출의 예[6]

먼저 군집화(Clustering)를 이용하여 심층신경망과 이진 나무(Binary Tree) 결정 구조를 결합하는 방법이 제안되었다[6]. 즉, 학습 데이터셋에 존재하지 않는 위변조 영상 패턴에 대해서 새로운 군집이 할당되고 이는 나무 구조의 각 노드를 통해 학습이 가능하도록 〈그림 3〉과 같은 방식으로 신경망을 구축하였다. 또한, 각 분기 노드에서는 잠재특징의 분별력이 극대화 될 수 있도록 투영(Projection) 기반 특징 분류를 수행하고 픽셀 단위의 위변조 영역 마스킹(Masking)을 통해 영역 별 분별력을 향상시키는 방법을 제안하였다. 다양한 실험 결과를 통해 실리콘 마스크, 부분 마스크 등 총 13 종류의 위변조 공격을 효과적으로 검출할 수 있음을 확인할 수 있다. Shao[7] 등은 도메인 일반화(Domain Generalization)를 기반으로 새로운 위변조 얼굴 영상에 강인한 심층신경망 구조를 제안하였다. 먼저, 도메인(즉, 위변조 형태)에 따라 특징을 추출할 수 있는 개별 신경망을 학습한다. 모든 도메인의 공통된 특징을 추출할 수 있는 신경망을 구축하여 개별 신경망의 분별자(Discriminator)가 이를 서로 다른 도메인 특징으로 인식할 수 없도록 적대 손실(GAN Loss)을 이용하여 학습한다. 또한, [3]의 방법과 마찬가지로 얼굴 깊이 영상을 이용하여 실제 얼굴 영상의 분별력을 향상 시킴으로써 학습 데이터에 존재하지 않는 유형의 위변조 얼굴 영상에도 강인한 동작이 가능하다.

다중 채널 입력 영상을 이용하여 얼굴 위변조 검출을 수행하는 연구 또한 진행되기 시작했다. 가장 대표적으로 CVPR 2019 학회에서 Liu[8] 등은 RGB, 근적외선 및 깊이 영상으로 이루어진 다중 채널 얼굴 영상을 기반으로 한 얼굴 위변조 검출 데이터셋을 공개하였다. 이를 바탕으로 고성능 얼굴 위변조 검출 방법들도 함께 제안되었으며, 대부분 특징 혼합(Fusion) 후 압축을 통해 다중 채널 간 얼굴 특징의 관계를 학습하는 구조의 심층신경망을 제안하였다. Shen[9] 등은 각 채널별 영상을 지역적 영역으로 분할하고 ResNet[10] 기반 인코더를 이용하여 특징을 추출한다. 채널별로 추출된 특징은 단순 결합(Concatenation) 후 합성곱 연산을 통해 압축되는 과정을 거쳐 위변조 점수 계산에 사용된다(〈그림 4〉 참조). Nikisins[11] 등은 다중 채널 입력 영상을 혼합하여 생성한 새로운 얼굴 영상(즉,



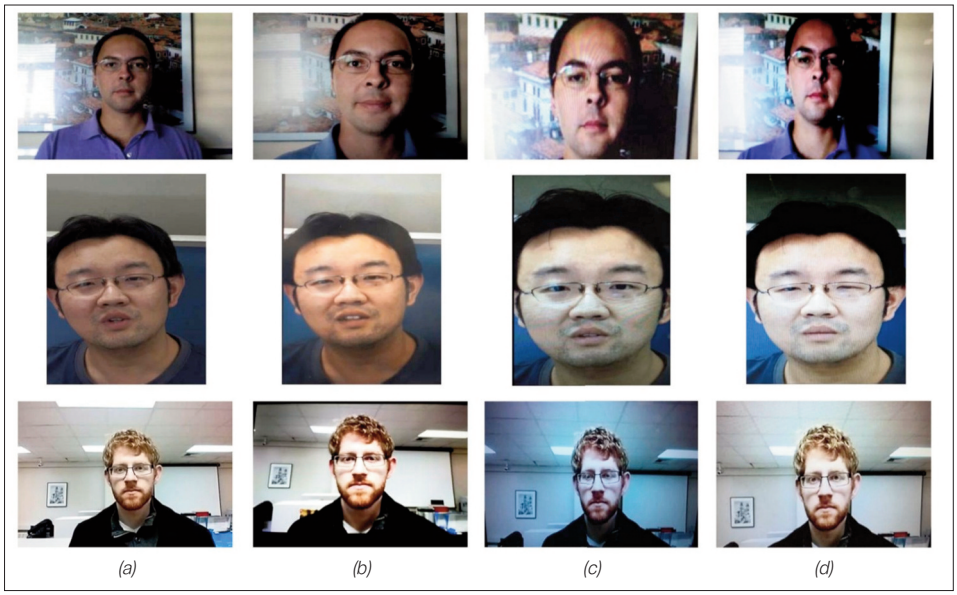
<그림 4> 다중 채널 입력 영상을 이용한 얼굴 위변조 검출의 예[9]

Multi-channel Face)을 학습에 사용하는 방법을 제안하였다. 인코더-디코더 구조에서 학습된 잠재 특징을 위변조 여부 판별에 사용하며, 이 때, 실제 얼굴만을 이용하여 학습을 수행한다. 이 밖에도 얼굴 영상 내 관심 영역을 학습하여 특징 가중치 재조정을 통해 검출 성능을 향상 시키는 방법[12], 동영상을 기반으로 얼굴 표정의 미세한 변화를 위변조 검출에 이용하는 방법[13], 위변조 얼굴 유형을 3D 모델을 이용하여 다변화 및 확장하여 다양한 형태를 효과적으로 학습할 수 있도록 하는 방법[14] 등이 제안되었다.

최근 연구 동향을 종합해 볼 때, 데이터 측면에서는 다중 채널 영상을 함께 이용하여 채널별 분별력 있는 특징을 효과적으로 학습할 수 있는 방법들이 연구되고 있고, 알고리즘 측면에서는 학습 데이터셋에 존재하지 않는 새로운 유형의 위변조 영상을 효과적으로 검출하기 위한 신경망 구조 및 학습 전략에 대한 연구가 활발히 진행되고 있다.

III. 얼굴 위변조 검출을 위한 데이터셋

이번 장에서는 얼굴 위변조 검출을 위해 사용되고 있는 데이터셋에 대해 간략히 살펴보고자 한다. 대부분 RGB 칼라 색상을 기반으로 하고 있으며, 사진 출력과 종이로 제작된 마스크 및 고해상도 디스플레이를 이용한 위변조 영상을 포함하고 있다. 가장 대표적으로 사용되고 있는 데이터셋으로는 CASIA-FASD[15], Replay-Attack[16]과 MSU-USSA[17] 데이터셋이 있다. CASIA-FASD와 Replay-Attack 데이터셋은 50명 규모의 피실험자로부터 영상을 획득하였으며, 위변조 영상 제작을 위해 iPad 및 프린트된 사진을 사용하였고, 실제 얼



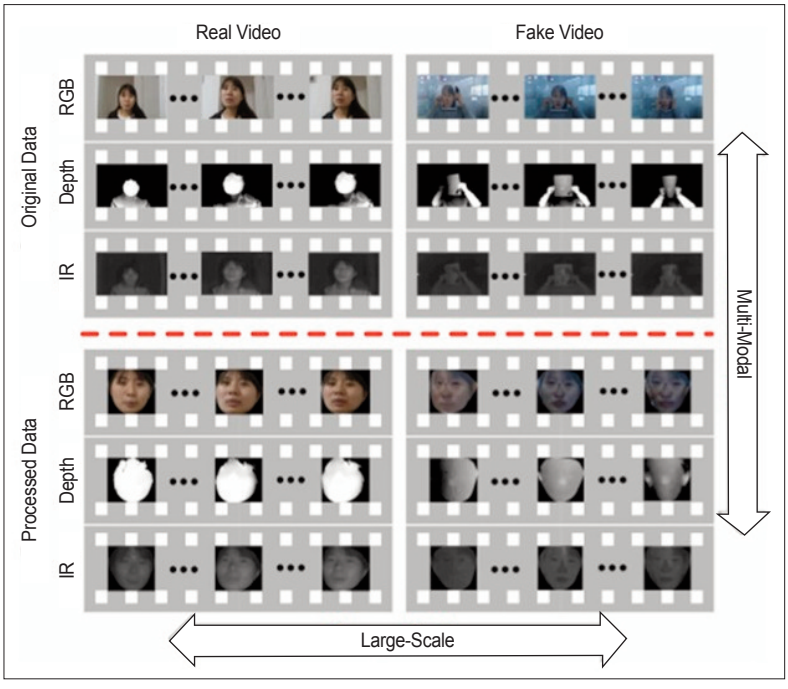
<그림 5> 얼굴 위변조 데이터셋의 예(Top : Replay-Attack / Middle : CASIA-FASD /Bottom : MSU-USSA) (a) 실제 얼굴, (b)-(d) 위변조 얼굴

굴은 MacBook 웹카메라, USB 카메라, 및 Sony Nex-5 모델을 사용하였다. MSU-USSA 데이터셋의 경우, 1,140명의 피실험자로부터 실제 얼굴 영상 및 위변조 얼굴 영상을 획득하였으며, 데이터셋 제작을 위해 다양한 기기(예를 들어, Nexus 5, Tablet 등)를 사용하였다. 〈그림 5〉는 각 데이터셋에 포함된 실제 얼굴 영상과 위변조 얼굴 영상의 예를 보여주고 있다.

가장 최근에는 다중 채널 입력 영상을 포함한 대용량 얼굴 위변조 검출 데이터셋(CASIA-SURF)이 발표되었다[8]. 총 1,000명의 피실험자로부터 21,000개의 비디오 샘플을 획득하였으며, RGB 칼라, 근적외선, 깊이 영상을 포함하고 있다. CASIA-SURF 데이터셋의 모든 영상은 Intel RealSense SR300 카메라를 이용하여 촬영되었으며, 학습, 검증(Validation) 및 테스트 서브셋으로 나누어 제공된다. 총 여섯 유형의 위변조 영상을 포함하고 있으며, 샘플 영상을 〈그림 6〉에 나타내었다. 이와 같은 대용량 데이터셋은 효과적인 심층신경망 학습을 가능하게 하여 얼굴 위변조 검출 성능을 크게 향상 시킬 수 있을 것으로 기대된다.

Ⅳ. 결 론

본 고에서는 얼굴 위변조 검출 기술, 특히, 심층학습을 이용한 얼굴 위변조 검출 최신 기술 동향에 대해 살펴보았다. 모바일 기기를 위한 생체 인증 기술, 특히, 얼굴 인증 기술이 꾸준히 발전하고 있지만, 다양한 위변조 공격으로 인해 고성능 보안 시스



<그림 6> 다중 채널 입력 영상을 이용한 얼굴 위변조 검출 데이터셋(CASIA-SURF)의 예[8]

템 구축에 여전히 어려움이 있다. 컴퓨터 비전 분야에서는 이를 해결하기 위해 최근 심층학습을 이용한 얼굴 위변조 검출 연구가 활발히 진행되고 있다. 특히, 학습 데이터셋에 존재하지 않는 위변조 공격 유형에 강인한 동작을 위해 생성 모델 및 군집화 등 다양한 기법을 심층신경망에 적용하고자 하는 연구가 진행되고 있으며, 학습 데이터도 기존 RGB 칼라만 사용하는 것에서 벗어나 근적외선 및 깊이 영상까지 활용하는 방법이 제안되고 있다. 학습 파라미터 경량화 및 위변조 공격 유형에 관계 없이 강인한 동작이 가능한 심층신경망 구조가 개발된다면 인공지능 기술 기반 얼굴 위변조 검출 방법이 모바일 기기에도 성공적으로 적용될 수 있을 것으로 기대된다.

참 고 문 헌

[1] Y. Atoum, Y. Liu, A. Jorabloo, and X. Liu, “Face anti-spoofing using patch and depth-based CNNs,” in Proc. IEEE Int. J. Conf. Biometrics, pp. 319-328, Oct. 2017.

[2] H. Li et al., “Learning generalized deep feature representation for face anti-spoofing,” IEEE Tr. Inform. Forensics and Secur., vol. 13, no. 10, pp. 2639-2652, Oct. 2018.

[3] Y. Liu, A. Jourabloo, and X. Liu, “Learning deep models for face anti-spoofing: binary or auxiliary supervision,” in Proc. IEEE CVPR, pp. 389-398, Jun. 2018.

[4] X. Yang, W. Luo, L. Bao, D. Gong, S. Zheng, Z. Li, and W. Liu, “Face anti-spoofing: model matters, so does data,” in Proc. IEEE CVPR, pp. 3507-3516, Jun. 2019.

[5] A. Jourabloo, Y. Liu, and X. Liu, “Face de-spoofing: anti-spoofing via noise modeling,” in Proc. European Conference on Computer Vision (ECCV), Sep. 2018.

[6] Y. Liu, J. Stehouwer, A. Jourabloo, and X. Liu, “Deep tree learning for zero-shot face anti-spoofing,” in Proc. IEEE CVPR, pp. 4675-4684, Jun. 2019.

[7] R. Shao, X. Lan, J. Li, and P. C. Yuen, “Multi-adversarial discriminative deep domain generalization for face presentation attack detection,” in Proc. IEEE CVPR, pp. 10015-10023, Jun. 2019.

[8] A. Liu et al., “Multi-modal face anti-spoofing attack detection challenge at CVPR 2019,” in Proc. IEEE CVPRW, Jun. 2019.

[9] T. Shen, Y. Huang, and Z. Tong, “FaceBagNet: bag-of-local-features model for multi-modal face anti-spoofing,” in Proc. IEEE CVPRW, Jun. 2019.

[10] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE CVPR, pp. 770-778, Jun. 2016.

[11] O. Nikisins, A. George, and S. Marcel, “Domain adaptation in multi-channel autoencoder based features for robust face anti-spoofing,” in Proc. ICB, pp. 1-8, Jun. 2019.

[12] H. Chen, Y. Chen, X. Tian, and R. Jiang, “A cascade face spoofing detector based on face anti-spoofing R-CNN and improved Retinex LBP,” IEEE Access, vol. 7, pp. 170116-133, Dec. 2019.

[13] X. Tu, H, Zhang, M. Xie, Y. Luo, Y. Zhang, and Z. Ma, “Enhance the motion cues for face anti-spoofing using CNN-LSTM architecture,” in arXiv 1901.05635, Jan. 2019.

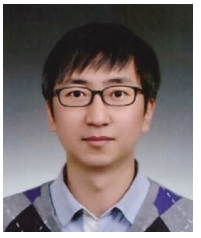
[14] J. Guo, X. Zhu, J. Xiao, Z. Lei, G. Wan, and S. T. Li, “Improving face anti-spoofing by 3D virtual synthesis,” in Proc. ICB, pp. 1-8, Jun. 2019.

[15] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z. Li, “A face antispoofing database with diverse attacks,” in Proc. ICB, 2012, pp. 26-31.

[16] I. Chingovska, A. Anjos, and S. Marcel, “On the effectiveness of local binary patterns in face anti-spoofing,” in Proc. IEEE BIOSIG, Sep. 2012, pp. 1-7.

[17] K. Patel, H. Han, and A. K. Jain, “Secure face unlock: spoof detection on smartphones,” IEEE Tr. Inform. Forensics and Secur., vol. 11, no. 10, pp. 2268-2283, Oct. 2016.

필 자 소 개



김 원 준

– 2012년 8월 : 한국과학기술원(KAIST) 박사
– 2012년 9월 ~ 2016년 2월 : 삼성종합기술원 전문연구원
– 2016년 3월 ~ 2020년 2월 : 건국대학교 전기전자공학부 조교수
– 2020년 3월 ~ 현재 : 건국대학교 전기전자공학부 부교수
– 주관심분야 : 컴퓨터 비전, 영상처리, 기계학습