Monthly Report
2026-5호

www.copyright.or.kr

저작권 기술 트렌드

COPYRIGHT TECH TREND

한국저작권위원회
KOREA COPYRIGHT COMMISSION

K-콘텐츠 불법 유통 대응 기술: 딥러닝 워터마킹과 디지털 핑거프린팅

뉴스 브리프

K-콘텐츠의 글로벌 인기와 함께 불법 유통도 빠르게 확산되고 있다. 불법 사이트는 차단된 즉시 새로운 도메인으로 이동하고, 콘텐츠는 클라우드·개인 간 파일 공유 등 추적이 어려운 경로로 유출되어 창작자와 투자자에게 정당한 보상이 돌아가기 어려운 구조가 형성된다. 이처럼 사후적 차단 조치가 한계를 드러내자, 콘텐츠 자체를 추적·식별하는 기술이 대안으로 부상하고 있다. 본 보고서는 영상 압축·변형 공격 환경에서 향상된 강건성을 보이는 딥러닝 비디오 워터마킹 기술 DINVMark와, 원본을 변경하지 않고 콘텐츠 고유의 특징을 추출하여 식별자를 생성하는 디지털 핑거프린팅을 다룬다. 두 기술은 콘텐츠 유출 시점의 추적과 유통 단계의 식별이라는 서로 다른 역할을 담당하며, 차세대 콘텐츠 보호 체계의 기반 기술로 발전할 가능성을 보여준다.

뉴스 플러스

Ⅰ. 서론 : 반복되는 저작권 분쟁과 AI 기반 대안의 모색

• K-콘텐츠의 글로벌 신드롬과 불법 유통의 그늘

K-콘텐츠는 전 세계적인 문화 현상으로 자리 잡았다. '오징어 게임'과 '폭싹 속았수다'와 같은 작품들은 글로벌 시청자를 사로잡으며 한국 대중문화 산업의 위상을 한 단계 끌어올렸다. 그러나 콘텐츠의 인기가 커질수록 불법 유통 문제도 함께 부각되고 있다. 실제로 2024년 한 해 동안 해외에서 불법 유통된 K-콘텐츠는 약 4억 1,400만 건으로 집계되었으며[1], 이는 산업 성장에 부담을 주는 요인으로 지목된다.

1) 한국저작권보호원, "'2024 해외 한류콘텐츠 침해 실태조사 보고서", 한국저작권보호원, 2025.01.31, https://www.kcopa.or.kr/lay1/bbs/S1T283C291/A/64/view.do?article_seq=6252&cpage=1&rows=10&condition=&keyword=

콘텐츠 불법 유통이 야기하는 핵심 문제는 창작 생태계의 선순환 구조에 악영향을 미친다는 점이다. 한 편의 드라마가 완성되기까지는 배우, 작가, 감독뿐만 아니라 수많은 스태프와 제작사, 투자사의 노력이 투입된다. 정당한 대가를 지불하고 콘텐츠를 소비하는 합법적 유통 구조는 이러한 노력이 다시 새로운 창작을 위한 투자로 이어지게 하는 동력이 된다. 그러나 불법 스트리밍과 무단 자막 유통은 이 과정을 우회하여 투자자의 정당한 수익 회수를 어렵게 만든다. 결국 콘텐츠 소비 규모는 확대되었지만 수익은 유출되는 기형적 구조가 자리 잡을 수 있으며, 이는 신규 콘텐츠에 대한 투자 위축과 창작 환경 악화로 이어져 장기적으로 K-콘텐츠의 제작 기반을 위협하는 결과를 초래할 수 있다.

특히 넷플릭스(Netflix) 등 글로벌 OTT 플랫폼이 공식적으로 서비스되지 않는 중국과 같은 지역에서는 상황이 더욱 두드러진다. 현지 리뷰 사이트인 더우반(豆瓣)에는 '월간남친', '흑백요리사', '오징어 게임 시즌3' 등 정식 출시되지 않은 K-콘텐츠의 리뷰와 평점이 게시되는 등 불법 시청이 폭넓게 이루어지고 있다.[2] 이는 저작권 침해에 그치지 않고, K-콘텐츠의 산업적 가치가 정당하게 평가받고 보상받을 기회를 제약하는 요인으로 작용한다.

• 도메인 우회와 차단·재생성의 반복

콘텐츠 불법 유통 문제에 대응하기 위해 정부와 업계는 지속적으로 노력을 기울여 왔다. 특히 2026년 5월 11일부터는 문화체육관광부 장관이 직접 정보통신망 사업자에게 불법 사이트 접속 차단을 명령할 수 있는 긴급차단제도가 시행되었으며, 시행 첫날 34곳의 불법 사이트에 대해 긴급 차단이 집행되는 등 행정적 대응이 한층 강화되었다.[3] 이러한 대응은 제도 시행 전 일부 불법 사이트의 운영 종료로 이어지는 성과를 보이기도 하였다. 다만 사후 차단 방식이 갖는 본질적 한계는 여전하다. 긴급차단제도 시행 이후에도 불법 사이트가 대체 사이트로 자동 연결되거나 텔레그램 등을 통해 새 주소를 안내하는 방식으로 차단을 우회하는 사례가 확인되고 있기 때문이다.[4]

이러한 양상은 중국 당국이 클라우드 저장 서비스 쿼크(Quark)를 단속한 사례에서도 확인된다.[5] 단속으로 기존 공유 링크가 대거 삭제되었으나, 이용자들이 대안 웹하드로 이동하며 불법적인 콘텐츠 소비를 이어갈 것이라는 전망이 나온다. 이는 플랫폼과 이용자, 단속 기관 간의 지속적인 대응 경쟁 구도가 지속될 수 있음을 보여준다.

불법 유통의 통로가 특정 웹사이트에 국한되지 않고 클라우드 저장소, 개인 간 파일 공유(P2P) 플랫폼, 커뮤니티 내 암호화된 링크 공유 등으로 점차 분산되고 파편화되는 흐름은 기존 대응 체계만으로는 충분하지 않음을 보여준다. 이는 저작권 보호의 접근 방식에서 콘텐츠 추적·식별 기술 도입의 필요성을 시사한다.

2) 김현덕, "'불법 시청 논란, K드라마 인기는 왜 해적판을 부르나", 스포츠서울, 2026.04.29, https://www.sportsseoul.com/news/read/1606239
3) 박정환, "뉴토끼 등 34곳 '우선 차단'…저작권법 개정 첫날부터 문체부 '본격대응'", 뉴스1, 2026.05.11, https://www.news1.kr/life-culture/general-cultural/6162669
4) 장병호, "'긴급차단' 우회하는 불법사이트…최휘영 장관, 전문가와 대책 논의", 이데일리, 2026.05.26, https://www.edaily.co.kr/News/Read?newsId=02528886645452856
5) 배인선, "[중국 화양'영'화] 'K드라마 불법유통 온상' 中 웹하드 단속…검열인가 저작권 보호인가", 아주경제, 2026.04.18, https://www.ajunews.com/view/20260417123914946

• **콘텐츠 보호 패러다임의 전환: 차단에서 추적으로**

기존의 불법 유통 대응 방식이 사이트 차단이라는 사후 조치에 집중되어 있었다면, 최근에는 콘텐츠 유출의 최초 근원지를 추적하는 방식으로 대응의 방향이 전환되고 있다. 불법 사이트의 주소는 계속 바뀔 수 있지만, 불법으로 유통되는 콘텐츠 자체는 변하지 않는다는 점에 착안한 접근이다. 즉 웹사이트 중심의 접근 방식이 아닌 콘텐츠 자체에 기반한 기술적 접근이 논의되고 있다.

이러한 변화의 중심에는 디지털 워터마킹(digital watermarking)과 디지털 핑거프린팅(digital fingerprinting) 기술이 있다. 디지털 워터마킹은 육안으로 식별하기 어려운 고유의 식별 정보*를 영상이나 음원을 포함한 모든 콘텐츠 자체에 삽입하는 기술이다. 해당 콘텐츠가 불법으로 유출될 경우, 삽입된 워터마크를 분석하여 유출 시점과 경로, 유출자에 관한 추적 단서를 확보할 수 있다. 디지털 핑거프린팅은 콘텐츠를 변경하지 않고 콘텐츠의 고유한 특징을 추출하여 별도의 식별자를 생성하는 기술이다. 생성된 식별자를 데이터베이스에 등록하여 유출 의심 콘텐츠와 대조함으로써, 불법 복제본을 식별하고 탐지하는 데 활용된다.

본 보고서는 딥러닝(deep learning)** 기반의 비디오 워터마킹과 디지털 핑거프린팅을 핵심 기술로 다룬다. 이어지는 본론에서는 두 기술의 작동 원리와 구현 방식, 다양한 공격 상황에서의 강건성(robustness) 수준을 살펴본다. 또한 실제 불법 유통 환경에서 발생할 수 있는 기술적 한계와 적용 시나리오를 검토함으로써, 차세대 콘텐츠 보호 전략 수립에 참고할 수 있는 기술적 기반을 정리하고자 한다.

* 식별 정보(identification information): 저작권자, 배포 경로, 타임스탬프 등 콘텐츠에 담기는 데이터의 내용 자체를 가리킴. 이와 구분하여 본 보고서에서 식별자(identifier)는 해당 정보를 특정 형식으로 인코딩한 결과값, 즉 비트열(bit sequence)이나 해시 값(hash value)과 같은 구체적인 값을 의미함.

** 딥러닝(deep learning): 여러 층으로 구성된 인공신경망을 이용해 대량의 데이터에서 특징과 패턴을 자동으로 학습하는 인공지능 기술. 영상 처리 분야에서는 프레임·픽셀·주파수 정보 등 복합적인 특징을 학습하여 워터마크 삽입·추출, 영상 인식, 품질 복원 등에 활용됨

II. 본론 1: 딥러닝 기반 비디오 워터마킹 기술의 원리와 구현

• **비디오 워터마킹의 기술적 요구사항: 비가시성, 강건성, 삽입 용량**

디지털 워터마킹 기술은 강건성, 비가시성(invisibility), 삽입 용량(capacity)이라는 세 가지 핵심 요구사항을 균형 있게 충족할 때 효과적인 저작권 보호 수단으로 기능한다. 세 요소는 어느 한쪽을 강조하면 다른 쪽이 약해지는 상충 관계에 있어, 기술적 난도를 높이는 주요 요인이다.

첫째, 강건성은 영상 압축, 포맷 변환, 화면 잘라내기(cropping), 흐림 효과(blur) 등 의도적이거나 비의도적인 변형 상황에서도 워터마크 정보가 유지되는 특성을 말한다. 둘째, 비가시성은 워터마크가 삽입된 이후에도 원본 콘텐츠의 시각적 품질이 유지되어, 시청자가 그 차이를 인지하기 어려운 특성을 의미한다. 마지막으로 삽입 용량은 콘텐츠에 담을 수 있는 워터마크 정보의 양을 의미한다. 유출자 식별 정보, 시간 정보 등 충분한 데이터를 담으려면 일정 수준 이상의 용량 확보가 필요하다.

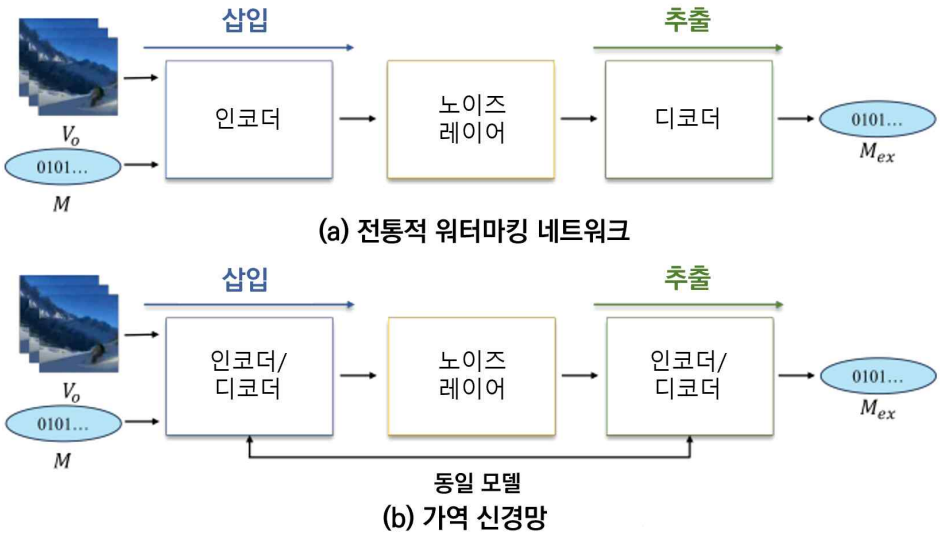
전통적인 워터마킹은 주로 영상의 특정 주파수 영역이나 픽셀 값에 정해진 규칙에 따라 정보를 삽입하는 방식으로 이루어졌다. 예를 들어 DT-CWT(dual-tree complex wavelet transform)*와 같이 영상을 주파수 대역 등으로 변환한 뒤 정보를 삽입하는 변환 영역 기반 기술은 영상의 특정 주파수 대역에 워터마크를 삽입하여 강건성을 높이도록 설계되었다. 다만 이러한 방식은 비가시성과 강건성, 삽입 용량 간의 균형을 확보하는 데 한계가 있었으며, 특히 고압축 영상 환경에서는 워터마크가 손상되기 쉬웠다.

* DT-CWT(dual-tree complex wavelet transform): 이중 트리 복소수 웨이블릿 변환. 영상이나 신호를 여러 주파수 대역으로 분해하여 분석하는 변환 기법으로, 영상의 방향성 정보를 잘 보존하고 변환 결과가 작은 변화에도 안정적이라는 특성이 있음. 디지털 워터마킹에서는 영상의 특정 주파수 대역에 워터마크를 삽입하여 외부 변형에 대한 강건성을 높이는 용도로 활용됨

- **DINVMark: 가역 신경망(INN) 기반 워터마킹 아키텍처**

최근에는 기존 워터마킹 기술의 한계를 보완하기 위해 딥러닝 기반 워터마킹 기술이 활발히 연구되고 있다. DINVMark(deep invertible network for video watermarking)는 그중에서도 대표적인 AI 기반 워터마킹 아키텍처로, 이 모델의 핵심은 가역 신경망(invertible neural network, INN)* 구조의 채택에 있다. 가역 신경망이란 입력 데이터를 처리하는 순방향(forward pass) 과정과 그 결과로부터 원본을 복원하는 역방향(backward pass) 과정이 하나의 네트워크 안에서 양방향 복원이 가능한 구조를 말한다.

[그림 1] 전통적 워터마킹 네트워크(a)와 가역 신경망 기반 DINVMark(b)의 구조 비교



출처: Jianbin Ji 외 4인, "DINVMark: A Deep Invertible Network for Video Watermarking", IEEE Transactions on Multimedia, 2025.09.22, https://arxiv.org/abs/2509.17416

전통적인 워터마킹 네트워크는 [그림 1]에서 볼 수 있듯이 워터마크를 삽입하는 인코더(encoder)와 추출하는 디코더(decoder)를 각각 별개의 모델로 설계한다. 반면 DINVMark는 가역 신경망의 특성을 활용하여 이 둘을 하나의 네트워크로 통합하였다. 순방향 연산이 삽입을, 역방향 연산이 추출을 담당하는 구조이다.

삽입과 추출이 동일한 네트워크 안에서 이루어지므로 정보 손실을 최소화할 수 있으며, 이는 워터마크 추출의 정확도를 높이는 데 기여한다.DINVMark의 또 다른 특징은 영상이 압축되는 과정에서 워터마크가 손상되는 영상 압축 공격(video compression attack)에 대한 강건성을 확보하기 위해 노이즈 레이어(noise layer)를 학습 과정에 도입한 점이다. 노이즈 레이어는 딥러닝 기반 워터마킹에서 일반적으로 사용되는 구성 요소이다. 학습 과정에서 워터마크가 삽입된 영상에 압축, 노이즈, 프레임 손실 등의 변형을 가하고, 네트워크가 이렇게 변형된 영상으로부터 워터마크를 정확히 추출하도록 훈련하는 역할을 한다.

기존 딥러닝 기반 워터마킹 모델은 H.264/AVC** 등 주로 구세대 코덱의 압축 환경을 위주로 설계되었으나, DINVMark는 고효율 비디오 코딩(HEVC, high efficiency video coding)*** 압축에 대응하도록 설계된 노이즈 레이어를 도입하였다는 점에서 구별된다. 현재 영상 유통 환경에서 HEVC가 널리 사용되고 있다는 점을 고려하면, 이는 실제 유통 환경에서의 강건성을 사전에 확보하려는 접근으로 볼 수 있다.

* 가역 신경망(invertible neural network, INN): 입력 데이터로부터 출력 데이터를 계산하는 순방향 과정과, 출력 데이터로부터 입력 데이터를 복원하는 역방향 과정을 모두 수행할 수 있는 신경망. 정보 손실이 적어 정밀한 데이터 변환 및 복원이 요구되는 분야에 활용됨

** H.264/AVC(advanced video coding): 영상 압축 표준의 하나로, 동영상을 효율적으로 압축하여 저장하거나 전송할 수 있도록 함. 2003년 표준화 이후 인터넷 영상, 방송, 블루레이 등 폭넓은 분야에서 사용되어 왔음

*** HEVC(high efficiency video coding, 고효율 비디오 코딩): H.264/AVC의 후속 영상 압축 표준으로, H.265라고도 불림. 동일한 화질을 기준으로 H.264/AVC 대비 약 절반 수준의 데이터 용량으로 영상을 압축할 수 있어, 4K·8K 등 고해상도 영상 전송에 널리 활용됨

• 딥러닝 기반 워터마킹 모델의 비교

AI 기반의 DINVMark가 등장하기 전까지 딥러닝 기반 워터마킹 분야에서는 여러 모델이 개발되었다. 이미지 워터마킹 분야에서 출발한 HiDDeN은 인코더-디코더 구조로 워터마크의 삽입과 추출을 학습하는 모델로 영상 워터마킹 모델의 기반이 되었다. 이후 등장한 DVMark는 인코더와 디코더에 다중스케일(multiscale)* 구조를 도입하여 영상의 시간적·공간적 차원에서 워터마크를 삽입할 수 있도록 설계되었다. 나아가 REVMark는 영상의 시간적 연관성을 학습하는 구조와 H.264/AVC 압축 시뮬레이터를 결합하여 영상 압축에 대한 강건성을 강화하였다. DINVMark는 이러한 흐름 위에서 가역 신경망 구조를 도입해 인코더-디코더 결합을 강화하고, 기존 모델이 다루지 않았던 HEVC 압축에 대응하는 노이즈 레이어를 적용한 모델이다. 해당 모델들의 특성을 비교하면 아래 [표 1]과 같다.

* 다중 스케일(multiscale): 영상이나 이미지를 여러 해상도·크기 수준에서 동시에 분석하는 방식. 작은 단위의 세부 특징과 전체적인 구조를 함께 포착할 수 있어, 영상 압축이나 변형에 대한 강건성을 높이는 기법으로 활용됨.

[표1] 딥러닝 기반 워터마킹 모델 비교

모델	삽입방식	비가시성	영상 압축 강건성	삽입 용량	파라미터 (M)	연산량 (G FLOPs)
HiDDeN	딥러닝 워터마킹	✓			0.53	35
DT-CWT	변환 영역 워터마킹	✓	✓		—	—
DVMark	딥러닝 워터마킹	✓	✓		21.04	2,361
REVMark	딥러닝 워터마킹	✓	✓		7.42	157
DINVMark	딥러닝 워터마킹	✓	✓	✓	1.38	54

* DT-CWT는 신경망 기반이 아니므로 파라미터·연산량 측정 대상에서 제외됨.

출처: Jianbin Ji 외 4인, "DINVMark: A Deep Invertible Network for Video Watermarking", IEEE Transactions on Multimedia, 2025.09.22, https://arxiv.org/abs/2509.17416 (논문 정보 기반 재구성)

• 성능 평가: 압축·변형 공격에 대한 강건성 검증

DINVMark의 성능은 다양한 영상 공격 시나리오를 대상으로 한 실험을 통해 검증되었다. 불법 스트리밍 환경에서 자주 발생하는 H.264/AVC 및 HEVC 영상 압축 공격에서, DINVMark는 기존의 딥러닝 기반 워터마킹 모델(HiDDeN, DVMark, REVMark 등)에 비해 높은 워터마크 추출 정확도를 보였다. 예를 들어 96비트 워터마크를 삽입한 환경(UCF-101* 데이터셋, HEVC QP=22 기준)에서 DINVMark는 약 99.98%의 비트 정확도(bit accuracy)를 기록하였다. 여기서 QP(quantizer parameter, 양자화 파라미터)는 영상 압축 과정에서 양자화 강도를 조절하는 값으로, 일반적으로 수치가 높을수록 압축률은 높아지지만 영상 품질은 낮아지고 워터마크 손상 가능성도 커진다. [그림 2] (d)에서 볼 수 있듯이, 압축 강도가 높아진 조건(QP=32)에서도 DINVMark는 약 90% 수준의 정확도를 유지하였으며, 같은 조건에서 REVMark는 약 72%, HiDDeN은 약 60% 수준에 머물렀다.

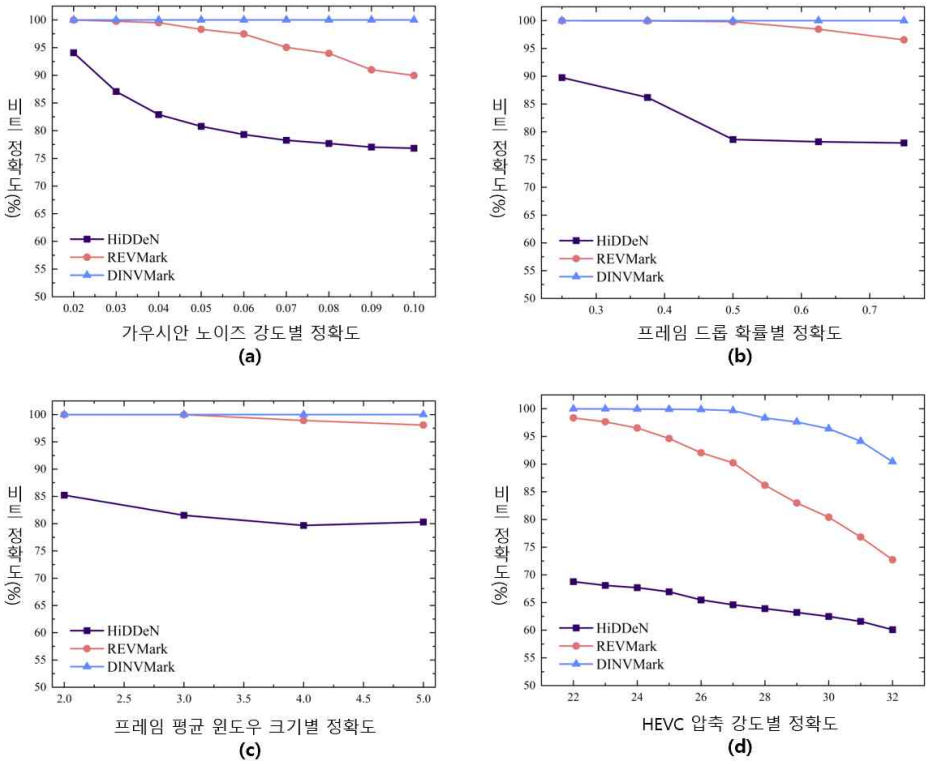
압축뿐만 아니라 (a) 가우시안 노이즈(Gaussian noise)**, (b) 프레임 드롭(frame drop)***, (c) 프레임 평균(frame average) 등 여러 유형의 변형 상황에서도 DINVMark는 100%의 추출 정확도를 보였다. 이는 가역 신경망 구조가 정보 손실을 줄이고, 학습 과정에 포함된 노이즈 레이어가 다양한 변형 상황에 대한 대응 능력을 강화한 결과로 해석할 수 있다.

* UCF-101: 미국 University of Central Florida에서 2012년 공개한 영상 데이터셋으로, 101개 행동 카테고리에 걸쳐 약 13,320개의 영상 클립으로 구성되며 영상 분석·인식 분야의 대표적인 벤치마크로 활용됨

** 가우시안 노이즈(Gaussian noise): 영상의 각 화소에 가우시안 분포(정규분포)를 따르는 무작위 잡음을 추가하여 변형을 가하는 방식. 영상 전송·압축 과정에서 흔히 발생하며, 워터마킹의 강건성 평가에 활용됨

*** 프레임 드롭(frame drop): 영상에서 일부 프레임을 임의로 제거하는 변형 방식. 일부 프레임이 손실된 상황에서도 워터마크가 안정적으로 추출되는지 평가하기 위한 강건성 테스트에 활용됨

[그림 2] 변형 공격 유형별 워터마크 추출 정확도 비교



출처: Jianbin Ji 외 4인, "DINVMark: A Deep Invertible Network for Video Watermarking", IEEE Transactions on Multimedia, 2025.09.22, https://arxiv.org/abs/2509.17416

기술적 강점은 삽입 용량에서도 확인된다. DINVMark는 96비트 수준의 정보를 안정적으로 삽입할 수 있을 뿐만 아니라, 일정한 구조적 패턴을 지닌 규칙형 워터마크를 기준으로 최대 1024비트까지 확장한 환경(HEVC QP=22)에서도 약 94%의 정확도를 유지하였다. 이는 유출자 정보, 타임스탬프, 배포 채널 식별자 등 다양한 추적 정보를 콘텐츠에 담을 수 있다는 점을 시사한다. 또한 DINVMark는 약 1.38M의 모델 파라미터*와 54G FLOPs**의 연산량을 보인다. 이는 DVMark, REVMark 등 기존 딥러닝 기반 모델에 비해 경량화(lightweight)된 구조이다. ([표 1] 참조)

다만 DINVMark에도 보완이 필요한 영역이 있다. 연구진은 결론부에서 영상의 일부를 잘라내는 크롭(cropping) 공격에 대해 충분한 강건성을 확보하지 못한 점, 트랜스코딩(transcoding)*** 공격 환경에서의 성능이 별도로 검증되지 않은 점을 한계로 명시하였다. 이는 워터마크가 공간 영역(spatial domain)에 삽입되는 구조적 특성에 기인하는 것으로, 향후 연구를 통해 보완이 필요한 영역이다.

* 모델 파라미터(model parameter): 딥러닝 모델이 학습 과정에서 조정하는 내부 변수. 파라미터 수가 많을수록 모델이 표현할 수 있는 정보량은 커지지만, 일반적으로 저장 공간과 연산 자원도 더 많이 필요함

** FLOPs(floating point operations): 딥러닝 모델이 영상을 처리하는 과정에서 수행하는 계산량을 나타내는 지표. 소수점이 포함된 숫자를 이용한 연산 횟수를 의미하며, 값이 클수록 일반적으로 더 많은 연산 자원과 처리 시간이 필요함

*** 트랜스코딩(transcoding): 동영상이나 음원 등 디지털 콘텐츠를 한 형식에서 다른 형식으로 변환하는 작업. 동일한 영상을 다양한 해상도(4K → HD)로 변환하거나, 한 압축 표준(H.264/AVC)으로 인코딩된 영상을 다른 압축 표준(HEVC)으로 다시 인코딩하는 과정 등이 해당함. 스트리밍 서비스에서 다양한 단말기와 네트워크 환경에 맞춰 콘텐츠를 제공하는 과정에서 널리 사용됨.

II. 본론 2: 디지털 핑거프린팅 기술의 메커니즘

• 디지털 핑거프린팅과 워터마킹의 기술적 구분

디지털 핑거프린팅은 디지털 워터마킹과 함께 콘텐츠 보호 분야에서 자주 언급되지만, 두 기술은 원본의 변경 여부와 활용 목적에서 차이를 보인다. 워터마킹은 식별 정보를 직접 삽입하여 콘텐츠의 일부를 변경하는 방식인 반면, 핑거프린팅은 원본을 변경하지 않고 고유의 특징만을 추출하여 별도의 식별자인 해시 값(hash value)*을 생성하는 방식이다.

[표2] 디지털 워터마킹과 디지털 핑거프린팅의 비교

구분	디지털 워터마킹	디지털 핑거프린팅
콘텐츠 변경 여부	식별자 삽입으로 일부 변경	변경 없이 원본 유지
식별자 위치	콘텐츠 내부	외부 데이터베이스
식별자 생성 방식	식별 정보를 콘텐츠에 직접 삽입	콘텐츠 고유 특징을 추출하여 해시 값 생성
주요 활용 목적	저작권자·유출자 추적	콘텐츠 식별·복제본 탐지

출처: Jianbin Ji 외 4인, "DINVMark: A Deep Invertible Network for Video Watermarking", IEEE Transactions on Multimedia, 2025.09.22., https://arxiv.org/abs/2509.17416
Wendi Chen 외 2인, "Digital Fingerprinting on Multimedia: A Survey", arXiv 2024.08.26, https://arxiv.org/abs/2408.14155
(논문 정보 기반 재구성)

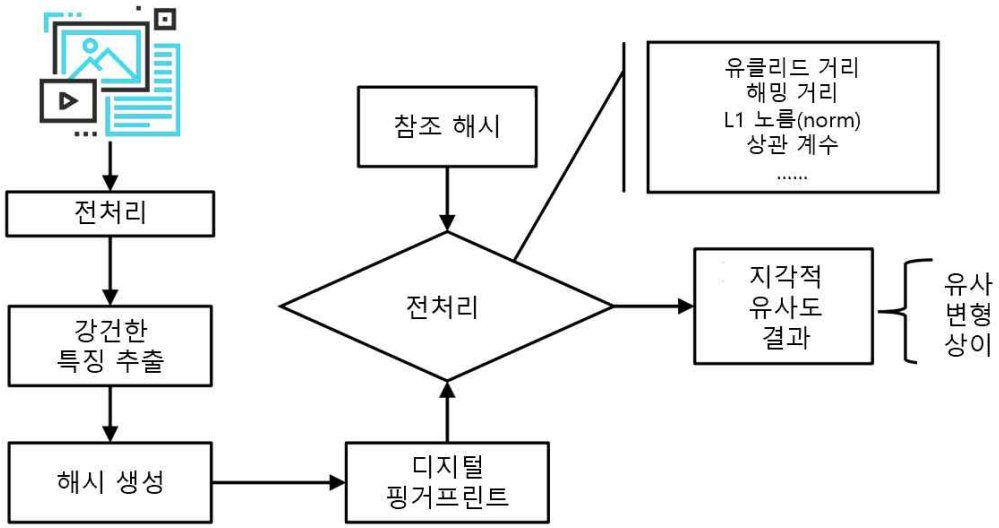
이러한 차이에 따라 활용 목적도 구분된다. 워터마킹은 콘텐츠에 삽입된 식별 정보를 통해 저작권자 정보나 유출자를 추적하는 데 활용되며, 본론 1에서 다룬 DINVMark 역시 이러한 워터마킹 기술에 해당한다. 반면 핑거프린팅은 콘텐츠 고유의 특징을 압축적으로 표현한 식별자를 데이터베이스에 별도로 저장한 뒤, 새로 발견된 콘텐츠가 기존 콘텐츠와 동일하거나 유사한지 판별하는 데 사용된다. 이처럼 두 기술은 콘텐츠 보호의 서로 다른 측면에서 상호 보완적으로 활용된다.

* 해시 값(hash value): 입력된 데이터를 일정한 규칙(해시 함수)에 따라 변환하여 얻은 짧은 길이의 결과 값. 동일한 입력에 대해서는 항상 동일한 결과가 산출되며, 데이터 검증·식별·비교 등에 활용됨. 본 보고서에서는 지각 해싱(perceptual hashing)을 통해 생성된 콘텐츠 식별자를 가리킴.

• 지각 해싱과 핑거프린트 생성 메커니즘

디지털 핑거프린팅의 핵심 기술은 지각 해싱(perceptual hashing)이다. 일반적인 암호 해시 함수는 입력 데이터에 미세한 변경만 가해져도 전혀 다른 해시 값을 출력하지만, 지각 해싱은 인간이 콘텐츠를 인지할 때 미세한 노이즈나 압축 손실을 무시하는 특성에 착안하여 콘텐츠의 핵심 특징만을 추출해 해시 값을 생성한다. 이에 따라 약간의 압축이나 변형이 가해져도 해시 값은 크게 변하지 않으며, 동일하거나 유사한 콘텐츠는 유사한 해시 값으로 표현된다.

[그림 3] 디지털 핑거프린팅 시스템 프레임워크



출처: Wendi Chen 외 2인, "Digital Fingerprinting on Multimedia: A Survey", arXiv 2024.08.26, https://arxiv.org/abs/2408.14155

핑거프린트 생성 과정은 [그림 3]의 좌측에서 볼 수 있듯이 일반적으로 세 단계로 구성된다. 먼저 콘텐츠의 크기 조정, 색 공간 변환 등 전처리 과정을 거친 뒤, 콘텐츠로부터 변형에 강건한 특징(색상 분포, 윤곽선, 주파수 영역 정보 등)을 추출한다. 이어 추출된 특징을 해시 함수를 통해 짧은 길이의 이진 시퀀스로 변환하여 해시 값을 생성한다. 이렇게 만들어진 해시 값이 곧 해당 콘텐츠의 디지털 핑거프린트가 된다.

• 핑거프린트 매칭 및 유사도 측정

생성된 핑거프린트는 콘텐츠 식별과 복제 탐지에 활용된다. 새로 발견된 콘텐츠가 데이터베이스에 등록된 원본과 동일하거나 유사한지 판별하는 과정은 [그림 3]의 우측에서 볼 수 있듯이 두 단계로 진행된다. 먼저 동일한 지각 해싱 알고리즘을 적용하여 의심 콘텐츠의 핑거프린트를 추출한다. 이어 추출된 핑거프린트를 데이터베이스에 저장된 참조 핑거프린트(reference fingerprint,)와 비교하여 일치 여부를 판단한다. 이때 데이터베이스의 핑거프린트는 [그림 3]에서 참조 해시로 표시된 값에 해당한다.

두 핑거프린트의 유사도는 거리 기반 지표로 측정된다. 대표적인 지표가 이진 형태의 두 해시 값에서 서로 다른 비트의 개수를 의미하는 해밍 거리(hamming distance)이다. 예를 들어 10110과 10011은 세 번째 비트(1→0)와 다섯 번째 비트(0→1)가 다르므로 해밍 거리는 2이다. 해밍 거리가 작을수록 유사도가 높은 핑거프린트로 판단할 수 있다. 이 외에도 두 점 사이의 직선 거리를 측정하는 유클리드 거리(euclidean distance), 각 원소 차이의 절댓값 합을 의미하는 L1 노름(L1 norm), 두 데이터의 패턴 유사도를 나타내는 상관 계수(correlation coefficient)등이 활용된다. 해시 값이 이진 형태일 경우 해밍 거리가, 정수나 실수 형태일 경우 유클리드 거리가 주로 사용되는 등 데이터 형태와 비교 목적에 따라 적합한 거리 측정 방식이 선택된다.

판별 결과는 미리 설정된 임계값을 기준으로 유사·변형·상이의 세 가지로 분류된다. 두 핑거프린트 간 거리가 임계값 미만일 경우 유사 콘텐츠로, 일정 오차 범위 내에 존재할 경우 변형 콘텐츠로, 기준을 초과할 경우 상이한 콘텐츠로 판단된다. 임계값 설정은 탐지 성능에 영향을 미친다. 임계값이 낮을 경우 약간의 변형만으로도 동일 콘텐츠를 놓치는 미탐(false negative)이 발생할 수 있고, 반대로 임계값이 높을 경우 서로 다른 콘텐츠를 동일 콘텐츠로 잘못 식별하는 오탐(false positive)이 발생할 수 있다. 예를 들어 대규모 불법 복제물 탐지 환경에서는 놓치는 사례를 줄이기 위해 임계값을 높게 설정할 수 있으나, 이 경우 오탐이 발생할 가능성이 높아진다. 이처럼 적용 환경의 목적과 허용 가능한 오류 수준에 따라 임계값을 적절히 설정하는 것이 핑거프린팅 시스템의 핵심 과제이다.

• 디지털 핑거프린팅의 기술적 도전 과제

디지털 핑거프린팅 기술은 콘텐츠 식별과 복제 탐지에 효과적으로 활용되고 있으나, 실제 적용 환경에서는 몇 가지 기술적 과제가 제기되고 있다.

첫째, 복합 공격(composite attacks)에 대한 대응이다. 실제 불법 유통 환경에서는 회전, 크기 조정, 블러, 노이즈 추가 등 여러 변형이 동시에 적용되는 경우가 많다. 기존 핑거프린팅 알고리즘은 단일 유형의 변형에는 비교적 안정적인 성능을 보이지만, 다수의 변형이 복합적으로 결합된 환경에서는 강건성이 저하될 수 있다.

둘째, 적대적 공격(adversarial attacks)이다. 적대적 공격은 사람의 눈으로 인지하기 어려운 수준의 미세한 수정을 콘텐츠에 가하여, 시각적으로는 원본과 동일하게 보이지만 핑거프린팅 시스템에서는 다른 해시 값이 산출되도록 유도하는 방식이다. 해당 공격으로 인해 불법 콘텐츠가 탐지를 회피한 채 유통될 가능성이 있어, 이를 방어하기 위한 알고리즘 강건성 확보가 주요 연구 과제이다.

셋째, 해시 역연산 공격(hash inversion attacks)이다. 지각 해싱은 콘텐츠의 핵심 특징을 해시 값에 담아내기 때문에, 해시 값으로부터 원본 콘텐츠의 일부 정보가 역으로 추정될 가능성이 있다. 이는 핑거프린팅 시스템 자체의 보안성과도 연결되는 과제에 해당한다.

넷째, 프라이버시 보호의 문제이다. 사용자 기기에서 콘텐츠를 사전에 스캔하여 핑거프린트를 비교하는 방식이 도입되는 사례가 늘면서, 이러한 스캔이 사용자 프라이버시에 영향을 미칠 수 있다는 우려가 나타나고 있다. 이에 핑거프린팅의 효율성과 프라이버시 보호 사이의 균형을 확보하기 위한 연구가 진행되고 있다.

이러한 과제는 디지털 핑거프린팅 기술이 단독 기술로 해결되기보다는, 본론 1에서 살펴본 워터마킹 기술과 상호 보완적으로 결합되는 방향에서 해소될 가능성이 있다. 두 기술의 조합은 콘텐츠 유출 시점의 추적(워터마킹)과 유통 단계의 식별(핑거프린팅)을 함께 지원함으로써, 다층적인 콘텐츠 보호 체계를 구성하는 핵심 기반 기술로 활용될 수 있다.

II. 본론 3: 불법 스트리밍 환경에서의 기술 적용 시나리오와 한계

• 실시간 스트리밍 환경에서의 워터마크 삽입 기술 과제

본론 1에서 살펴본 DINVMark와 같은 딥러닝 기반 워터마킹 기술은 파일 기반의 VOD(주문형 비디오) 환경에서 안정적인 성능을 보였다. 다만 이를 라이브 방송과 같은 실시간 스트리밍 환경에 그대로 적용하기에는 몇 가지 기술적 과제가 제기된다. 가장 두드러지는 과제는 처리 지연(latency) 문제이다. 실시간 스트리밍은 인코딩, 전송, 디코딩 과정이 수 초 이내에 이루어지는 구조이므로, 워터마크 삽입 과정에서 발생하는 추가 연산이 전체 파이프라인에 병목으로 작용할 경우, 영상 끊김이나 송출 지연이 발생할 수 있으며, 이는 라이브 방송의 실시간성과 시청자의 몰입을 저해하는 요인이 될 수 있다.

이러한 점에서 워터마킹 모델의 경량화는 실시간 스트리밍 환경의 지연 한계를 극복하기 위한 핵심 기술적 과제이다. DINVMark는 약 1.38M의 모델 파라미터와 54G FLOPs의 연산량으로 비교 모델 대비 경량화된 구조를 갖추고 있어, 실시간 환경 적용 가능성을 높이는 요소로 평가된다. 다만 라이브 스트리밍 환경에서는 네트워크 상황에 따라 전송률이 유동적으로 변하는 가변 비트레이트(adaptive bitrate)* 특성이 있다. 이로 인해 다양한 화질과 압축률 변화에서도 워터마크 검출 안정성을 확보하는 것이 추가 과제로 제기된다.

* 가변 비트레이트(adaptive bitrate): 네트워크 대역폭 변화에 따라 영상 화질과 전송률을 실시간으로 조정하는 스트리밍 방식.

• 클라우드 저장소·웹하드 유통 경로 추적의 실효성

최근 K-콘텐츠의 주요 불법 유통 통로로 지목되는 곳은 중국의 쿼크(Quark)와 같은 클라우드 저장 서비스(웹하드)이다. 이용자들은 비공개 커뮤니티와 메신저를 통해 암호화된 공유 링크를 주고받으며, 이는 전통적인 불법 스트리밍 사이트보다 추적을 한층 어렵게 만든다. 이러한 환경에서 워터마킹과 핑거프린팅을 적용하는 시나리오는 이론적으로 가능하지만 현실적인 제약이 따른다.

워터마킹 기술을 적용하면 콘텐츠를 배포하는 시점에 사용자별로 서로 다른 식별 정보를 영상에 삽입해 둘 수 있다. 이후 클라우드에 유출된 영상이 확보되면, 삽입된 워터마크를 검출하여 최초 유출자를 특정하는 것이 가능하다. 다만 클라우드 저장소에 업로드된 파일을 외부에서 확보하는 과정 자체가 용이하지 않다. 일부 서비스는 파일을 여러 조각으로 분산 저장하거나 독자적인 암호화 방식을 적용하고 있어, 외부에서의 파일 분석이 제한되는 경우가 있다.

핑거프린팅 기술도 클라우드 환경에 적용할 수 있다. 원본 콘텐츠의 핑거프린트를 데이터베이스에 사전 등록해두면, 클라우드에서 수집된 의심 파일과 대조하여 불법 복제본을 식별할 수 있다. 다만 비공개 채널을 통해 은밀하게 유통되는 공유 링크를 수집하고, 유통된 파일에서 핑거프린트를 추출하는 작업은 적지 않은 시간과 자원이 요구된다.

결국 두 기술이 클라우드·웹하드 환경에서 마주하는 공통 과제는 탐지 대상 파일의 확보에 있다. 워터마킹이나 핑거프린팅이 이론적으로 유출자를 특정하거나 복제본을 식별할 수 있더라도, 분석 대상이 되는 파일에 접근하기 어려운 환경에서는 기술의 실효성이 제한될 수 있다.

• 자막 파일·클립 영상 등 파생 콘텐츠 추적의 한계

K-콘텐츠의 불법 유통은 원본 영상의 단순 배포에 그치지 않고, 다양한 형태의 파생 콘텐츠로 빠르게 확산되는 양상을 보인다. 한 OTT 업계 관계자는 "K-콘텐츠는 공개 직후 번역 자막과 클립이 빠르게 퍼진다"고 밝힌 바 있다.[6] 이러한 파생 콘텐츠는 현재의 워터마킹 및 핑거프린팅 기술이 충분히 대응하기 어려운 영역이다.

먼저 자막 파일(.smi, .srt)은 영상 콘텐츠와는 별개로 유통되는 텍스트 파일이다. 비디오 워터마킹은 영상의 픽셀 값에 정보를 삽입하고, 영상 핑거프린팅 역시 영상의 시각적 특징을 추출하는 방식이므로, 두 기술 모두 텍스트 형태의 자막 파일에는 직접 적용하기 어렵다. 비공식 자막 공유 커뮤니티를 통해 자막만 별도로 배포되는 경우, 영상 출처를 추적하는 작업은 제한될 수 있다.

짧은 클립 영상이나 움짤(GIF, WebP)로 불리는 짧은 반복 영상도 과제로 남는다. DINVMark와 같은 비디오 워터마킹 기술은 영상의 시간적·공간적 특징을 활용하여 워터마크를 삽입한다. DINVMark의 경우 영상의 일부를 잘라내는 크롭(cropping) 공격에 대해 충분한 강건성을 확보하지 못하는 한계가 있어, 전체 영상에서 극히 일부 구간만 남는 클립 영상에서는 유의미한 워터마크 정보를 복원하기 어려울 수 있다. 핑거프린팅 역시 콘텐츠의 전체적 특징을 기반으로 식별자를 생성하기 때문에, 원본의 일부 구간만 포함된 짧은 클립에서는 식별 정확도가 떨어질 가능성이 있다. 이처럼 불법 유통의 파편화 양상은 워터마킹과 핑거프린팅 두 기술 모두에 새로운 과제를 제기한다.

6) 김현덕, "'불법 시청 논란, K드라마 인기는 왜 해적판을 부르나", 스포츠서울, 2026.04.29, https://www.sportsseoul.com/news/read/1606239

• **콘텐츠 보호 대응 체계에서의 기술 운용 과제**

최근에는 AI를 활용한 저작권 보호 대응 체계의 고도화가 추진되는 흐름이 나타나고 있다. 워터마킹과 핑거프린팅 기술이 이러한 대응 체계에 활용되려면 대규모 운용 환경에 맞는 기술적 보완이 필요하다.

워터마킹 기술의 경우, 본론 1에서 살펴본 바와 같이 DINVMark는 비교 모델 대비 경량화된 구조를 갖추고 있다. 그러나 대규모 콘텐츠에 대해 사용자별 워터마크를 실시간으로 삽입하고 유출 시 이를 자동으로 검출하는 전 과정을 하나의 시스템 안에서 처리하려면, 연산 자원의 배분과 처리 속도 간의 균형이 과제로 남는다.

핑거프린팅 기술의 경우, 대규모 콘텐츠를 대상으로 자동 탐지를 수행하려면 방대한 양의 핑거프린트를 실시간으로 대조하는 처리 능력이 요구된다. 본론 2에서 살펴본 것처럼 임계값 설정에 따라 미탐과 오탐의 균형이 달라지는데, 대규모 환경에서는 이러한 오류가 누적될 가능성이 높아 임계값 최적화와 예외 처리 설계가 시스템 신뢰도를 좌우하는 요소이다. 이처럼 워터마킹과 핑거프린팅 기술이 실제 콘텐츠 보호 체계의 구성 요소로 자리 잡기 위해서는, 개별 기술의 성능뿐 아니라 대규모 운용 환경에서의 안정성과 정확성을 동시에 확보할 필요가 있다.

Ⅲ. 결론 및 전망

• **현재 기술 수준의 종합 평가: 강건성과 실용성의 간극**

본 보고서에서 살펴본 딥러닝 기반 비디오 워터마킹과 디지털 핑거프린팅 기술은 K-콘텐츠 불법 유통 문제에 대응하기 위한 새로운 기술적 접근을 제시한다. DINVMark와 같은 최신 워터마킹 기술은 가역 신경망(INN) 구조와 HEVC 압축에 대응하는 노이즈 레이어를 활용하여, 영상 압축 및 변형 공격 환경에서 한층 향상된 강건성을 보인다. 이러한 접근은 사후 차단 방식의 한계를 보완하면서, 콘텐츠 유출의 출처를 직접 추적할 수 있는 가능성을 보여준다. 디지털 핑거프린팅 역시 콘텐츠 고유의 특징을 추출하여 식별자를 생성하는 방식으로, 대규모 콘텐츠 데이터베이스에서 불법 복제본을 탐지하는 기술적 수단을 제공한다.

다만 이러한 기술적 성과가 실제 불법 스트리밍 환경에서 곧바로 실효성을 발휘하기에는 보완할 부분이 남아 있다. 복합 공격(composite attacks), 적대적 공격(adversarial attacks), 해시 역연산 공격(hash inversion attacks) 등 고도화된 무력화 시도에 대한 방어 기술은 연구 단계에 있다. 또한 자막 파일이나 짧은 클립 영상과 같이 파편화된 형태의 파생 콘텐츠의 경우 기술적 대응력이 제한적이며, 관련 연구가 진행 중이다. 결국 현재의 기술 수준은 통제된 실험 환경에서의 가능성을 보여주고 있으나, 다변화된 실제 유통 환경 전반을 포괄하기까지는 추가적인 연구·개발이 필요한 단계이다.

• **향후 기술 고도화 방향: 경량화, 멀티모달, 양자내성**

향후 워터마킹 및 핑거프린팅 기술은 실용성을 높이는 방향으로 발전할 것으로 전망된다.

첫째, 모델 경량화는 실시간 처리가 요구되는 라이브 스트리밍 및 온라인 동영상 플랫폼(OTT) 서비스에 기술을 적용하기 위한 필수 전제 조건이다. 연산 효율을 높여 서버 부하와 처리 지연을 줄이는 방향으로 연구가 이어질 가능성이 있다.

둘째, 멀티모달(multimodal) 워터마킹 기술도 논의되고 있다. 기존 워터마킹이 영상이나 음성 중 한 가지 미디어에만 워터마크를 삽입하는 단일 모달 방식이었다면, 멀티모달 워터마킹은 영상과 음성 등 여러 미디어에 동시에 워터마크를 삽입하는 방식이다. 영상이 심하게 변형되더라도 오디오 트랙에 삽입된 워터마크를 통해 유출 정보가 복원될 수 있다는 점에서 단일 모달 방식보다 강건성 측면에서 보완적 특성을 갖는다.

마지막으로 장기적 관점에서는 양자 컴퓨팅 환경에 대비한 양자내성(quantum-resistant)* 암호 알고리즘을 워터마킹 및 핑거프린팅에 결합하려는 연구도 진행될 수 있다. 이는 미래의 보안 환경 변화에 대응하기 위한 사전적 접근에 해당한다.

* 양자내성(quantum-resistant): 양자 컴퓨터의 강력한 연산 능력으로도 해독이 어렵도록 설계된 암호 알고리즘의 특성. 현재 널리 사용되는 RSA·ECC 등 공개키 암호는 양자 컴퓨터가 본격 실용화될 경우 안전성이 위협받을 가능성이 있어, 이에 대응하는 알고리즘 연구가 진행되고 있음

• **기술 융합을 통한 차세대 콘텐츠 보호 생태계 전망**

워터마킹과 핑거프린팅 기술은 단독으로 활용되기보다는 다른 기술과 결합하여 차세대 콘텐츠 보호 생태계의 구성 요소로 발전할 것으로 전망된다. 대표적인 사례로 블록체인(blockchain) 기술과의 결합을 들 수 있다. 콘텐츠의 핑거프린트를 블록체인에 기록해두면, 거래 단계마다 등록된 핑거프린트와 대조하여 유통 경로를 추적할 수 있다. 블록체인의 위·변조가 어려운 특성과 핑거프린팅의 콘텐츠 식별 특성이 결합되면, 콘텐츠 유통의 투명성을 높이는 방향으로 활용될 여지가 있다.

한편 생성형 AI의 확산은 AI 산출물의 진위와 출처를 확인해야 할 필요성을 새롭게 제기하고 있다. 워터마킹과 핑거프린팅 기술은 불법 복제물 추적이라는 본래의 역할을 넘어, AI가 생성한 합성 콘텐츠(synthetic content)의 출처를 인증하거나 딥페이크(deepfake)와 같은 위·변조 콘텐츠의 확산에 대응하는 기술로 확장될 가능성이 있다. 이러한 융합이 실효성을 갖추기 위해서는 기술 표준화를 위한 산업계의 협력과 국제적인 기술 협력 체계가 함께 갖추어져야 할 과제로 남아 있다.

참고문헌

- Jianbin Ji 외 4인, "DINVMark: A Deep Invertible Network for Video Watermarking", IEEE Transactions on Multimedia, 2025.09.22, https://arxiv.org/abs/2509.17416
- Wendi Chen 외 2인, "Digital Fingerprinting on Multimedia: A Survey", arXiv 2024.08.26, https://arxiv.org/abs/2408.14155
- 한국저작권보호원, "'2024 해외 한류콘텐츠 침해 실태조사 보고서", 한국저작권보호원, 2025.01.31, https://www.kcopa.or.kr/lay1/bbs/S1T283C291/A/64/view.do?article_seq=6252&cpage=1&rows=10&condition=&keyword=
- 김현덕, "'불법 시청 논란, K드라마 인기는 왜 해적판을 부르나“, 스포츠서울, 2026.04.29, https://www.sportsseoul.com/news/read/1606239
- 박정환, "뉴토끼 등 34곳 '우선 차단'…저작권법 개정 첫날부터 문체부 '본격대응'", 뉴스1, 2026.05.11, https://www.news1.kr/life-culture/general-cultural/6162669
- 장병호, "'긴급차단' 우회하는 불법사이트…최휘영 장관, 전문가와 대책 논의", 이데일리, https://www.edaily.co.kr/News/Read?newsId=02528886645452856
- 배인선, "[중국 화양'영'화] 'K드라마 불법유통 온상' 中 웹하드 단속…검열인가 저작권 보호인가", 아주경제, 2026.04.18, https://www.ajunews.com/view/20260417123914946