

저작권 기술 트렌드



COPYRIGHT TECH TREND



한국저작권위원회
KOREA COPYRIGHT COMMISSION

불법 유통 대응의 공백과 워터마크를 통한 유출 출처 추적

뉴스 브리프

2026년 7월 17일 개봉한 크리스토퍼 놀란 감독의 영화 ‘오디세이’가 개봉한지 일주일여 만에 불법 복제본이 소셜 플랫폼 X에 업로드되는 사건이 벌어졌다. 해당 게시물은 삭제되기까지 약 두 시간 반 동안 210만 회가 넘는 조회 수를 기록했다. 배급사 유니버설 픽처스는 즉시 삭제 절차를 진행했고 게시 계정도 정지되었으나, 이튿날 또 다른 판본이 새로 올라왔다. 영화가 흥행을 이어 간 것과 별개로, 이 사건은 침해물을 찾아 지우는 사후 대응만으로는 유출을 근본적으로 해결할 수 없음을 보여 준다. 콘텐츠에 식별 정보를 미리 새겨 두고 유통된 사본에서 이를 읽어내는 워터마크 기술은 이 공백을 메우려는 시도다. 워터마크가 답할 수 있는 단계별 질문 유형과 워터마크에 지불해야 하는 대가를 규명한 이론 연구와, 화면 재촬영 환경에서 유출 지점을 지목하는 기술 연구를 본 보고서에서는 함께 살펴본다.

뉴스 플러스

1. 서론: 콘텐츠 불법 유통과 유출 출처 추적의 필요성

• ‘오디세이’ 불법 복제본의 온라인 확산

호메로스의 서사시를 원작으로 삼은 크리스토퍼 놀란(Christopher Nolan) 감독의 영화 ‘오디세이(The Odyssey)’는 2026년 7월 17일 북미에서 개봉해 흥행을 이어 갔으나, 개봉 일주일을 조금 넘긴 시점부터 소셜 플랫폼 엑스(X)에서 불법 복제본이 유통되기 시작했다. 영화 전문 매체 할리우드 리포터(The Hollywood Reporter)는 ‘오디세이’와 같이 기대를 모은大作에서 불법 복제본이 나오는 일은 충분히 예상되는 사안이며, 스튜디오도 광범위 개봉에 앞서 이에 대비하는 경우가 많다고 전했다.

7월 25일에는 비교적 화질이 좋은 판본이 엑스에서 유통되었으며, 이 게시물은 삭제되기 전까지 수백만 회의 조회 수와 수천 건의 리포스트를 기록했고 게시한 계정은 정지되었다. 확산 과정에서 한 이용자는 누군가 ‘오디세이’ 전편을 엑스에 올렸다는 내용의 글을 남기며 해당 복제본을 다시 공유하기도 했다. 한편, 영화 전문 매체 버라이어티(Variety)는 유출된 영상이 고화질 판본임을 확인하고, 해당 게시물이 약 두 시간 반 만에 210만 회를 넘는 조회 수를 기록한 뒤 저작권 안내로 대체되었다고 보도했다.

이튿날인 7월 26일 오전에는 저화질 판본이 새로 게시되었다. 이 게시물 역시 약 두 시간 뒤 삭제되었고, 그때까지 5만 회에 가까운 조회 수를 기록했다. 삭제된 자리에는 저작권자의 신고에 따라 해당 미디어가 비활성화되었다는 안내 문구가 표시되었다. 하루 사이에 서로 다른 두 판본이 잇따라 올라온 셈이다.

• 삭제 중심 대응의 한계

할리우드 리포터에 따르면 스튜디오는 개별 링크와 게시물을 일일이 추적하는 작업과 함께, 더 넓은 생태계에서 벌어지는 대규모 불법 유통에도 대응한다. 즉, 침해 대응은 개별 게시물 단위와 유통망 단위의 두 층위에서 이루어진다. ‘오디세이’ 사안에서 확인된 조치는 그중 앞의 층위에 해당하는 것으로, 문제가 된 게시물을 저작권 안내로 대체하고 이를 올린 계정을 정지시키는 방식이었다. 배급사인 유니버설 픽처스(Universal Pictures)는 무단 게시를 인지한 즉시 삭제 절차를 가동했다고 밝히면서, 자사 콘텐츠와 지식재산권을 보호하기 위해 모든 적절한 구제 수단을 강구하겠다는 입장을 내놓았다.

그러나 이러한 조치는 침해가 발생한 뒤에 작동한다는 점에서 시간상의 제약을 떠안게 된다. 고화질 판본을 담은 게시물은 저작권 안내로 대체되기 전까지 210만 회를 넘는 조회 수를 기록했으며, 약 두 시간 반이 걸렸다. 이튿날 게시된 저화질 판본 역시 약 두 시간 만에 삭제되었으나 5만 회에 가까운 조회 수를 남겼다. 두 사례 모두 삭제까지 걸린 시간은 두 시간 안팎에 그쳤지만, 그사이 상당한 규모의 확산이 이루어진 셈이다. 나아가 하루 사이에 화질이 다른 두 판본이 잇따라 게시되었다는 사실은, 개별 게시물에 대한 조치가 동일 저작물의 유통 자체를 멈추지는 못했음을 보여 준다.

대응에 걸리는 시간이 저작권자의 조치만으로 결정되지 않는다는 점도 확인된다. 할리우드 리포터는 소셜 플랫폼에 올라온 불법 복제물이 삭제되기까지 걸리는 시간이 기업마다 다르다고 전했다. 특히, 엑스는 이러한 요청에 협조를 얻기 어려운 곳으로 알려져 있다는 것이다. 저작권자가 같은 절차를 밟더라도 어느 플랫폼에 게시되었는지에 따라 실제 처리 결과가 달라질 수 있다는 의미다.

다행스럽게도 이번 유출에도 불구하고 ‘오디세이’는 높은 상업적 성과를 보였다. ‘오디세이’는 개봉 2주차에 9,000만 달러(약 1,236억 원)¹⁾를 벌어들이며 개봉 10일 만에 전 세계 6억 5,200만 달러(약 8,956억 원)를 기록했다. 다만, 이러한 흥행 지표에도 불구하고 유통된 복제본이 어디에서 비롯되었는지에 관해서는 아무것도 알려진 바가 없으며, 유출 경로에 관한 내용 역시 보도된 바 없다.

1) 1달러 = 1,373.60원(2026.09.01. KEB 하나은행 매매기준율 적용), 이하 동일 기준 적용

• 출처 규명 수단으로서의 워터마크

앞서 살펴본 대응은 이미 유통된 침해물을 정리하는 데 초점이 있다. 게시물을 내리고 계정을 정지시키는 조치는 확산을 멈추는 데는 유효하지만, 그 복제본이 어느 단계에서 만들어졌는지에 대해서는 어떠한 단서도 주지 못한다. 침해물을 발견해 제거하는 일과 그 침해물의 출처를 규명하는 일은 서로 다른 성격의 문제이며, 전자를 아무리 신속하게 수행해도 후자가 자동으로 해결되지는 않는다. 유출이 반복되는 상황에서 대응의 실효를 높이려면, 결과를 지우는 절차와 별개로 출처를 확인할 수 있는 수단이 필요하다.

이러한 요구에 대응하는 기술적 방식 가운데 하나가 워터마크(watermark)다. 워터마크란 사람이 알아채기 어려운 형태의 정보를 콘텐츠에 미리 삽입하고, 이후 유통된 사본에서 그 정보를 읽어내는 기술을 의미한다. 워터마크에는 콘텐츠 출처에 관한 단서가 있으므로, 워터마크를 이용하면 침해물이 발견되었을 때 그 사본이 어느 경로를 거쳐 나왔는지 확인할 근거가 마련된다. 다만, 워터마크가 삽입되어 있다고 해서 모든 형태의 질문에 답할 수 있는 것은 아니다. 워터마크가 존재하는지를 판정하는 일과 그것이 누구의 것인지를 지목하는 일은 서로 다른 문제이며, 각각에 필요한 조건도 역시 다르다.

본 보고서는 이 문제를 서로 다른 각도에서 다룬 두 건의 연구를 살펴본다. 첫 번째 연구는 워터마크가 답할 수 있는 질문을 네 단계로 나누고, 각 단계에 어느 정도의 정보량이 필요한지를 정보이론의 관점에서 규명한다. 두 번째 연구는 화면에 표시된 콘텐츠를 카메라로 되찍는 재촬영 환경에서 유출 지점을 지목하는 방식을 제안하고, 그 과정에서 새롭게 드러나는 공격 유형과 대응책을 함께 제시한다. 두 연구 중 하나는 이론적 한계선을 다루고 다른 하나는 구체적 구현을 다룬다는 점에서 이들 두 연구의 접근 방법은 다르다. 하지만 두 연구 모두 유통된 사본에서 출처를 읽어내려면 필요한 것이 무엇인지 보여 준다. 이어지는 본론에서는 각 연구의 원리와 검증 결과를 차례로 분석하고, 두 연구가 만나는 지점을 짚어 본다.

II. 본론 1: 워터마크의 포렌식 기능과 정보이론적 한계

• 워터마크로 수행할 수 있는 네 가지 과제

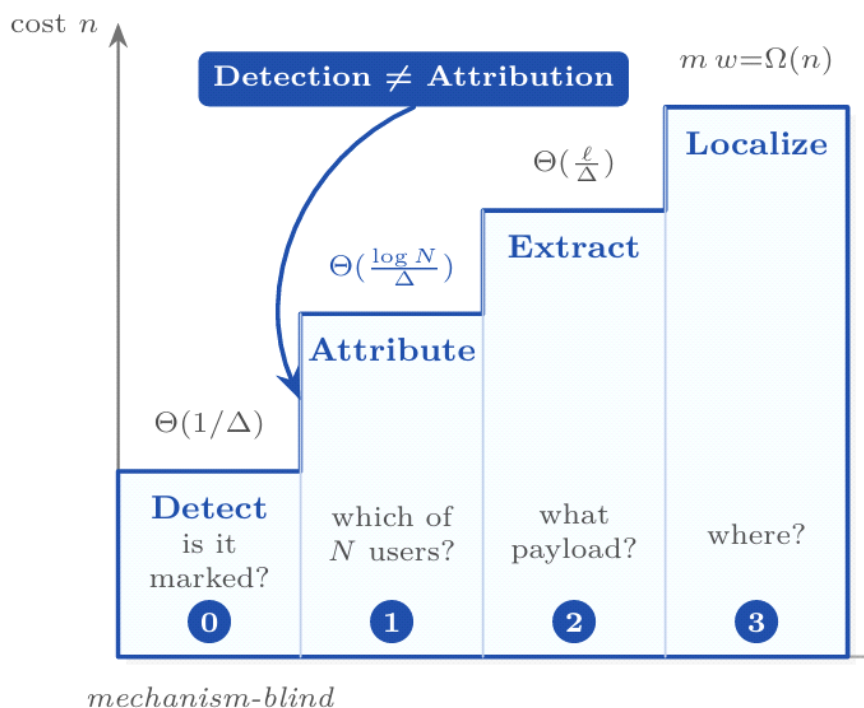
호주 뉴사우스웨일스대학교(University of New South Wales)와 그리피스대학교(Griffith University)가 2026년 7월 공개한 연구는 생성형 모델의 결과물에 삽입된 워터마크를 정보이론의 관점에서 분석한다. 워터마크를 만드는 방법이 아니라, 워터마크가 남긴 정보의 양으로 무엇을 할 수 있는가를 따지는 접근이다. 지금까지 워터마크에 관한 논의는 워터마크의 존재 여부를 가려내는 문제에 집중된 경향이 있다. 그러나 워터마크라는 같은 장치를 이용하여 그 이상의 과제를 수행할 수도 있는데, 이 연구는 그 과제를 네 가지 유형으로 구분한다.

첫 번째 과제는 탐지(detection)다. 대상 콘텐츠에 워터마크가 삽입되어 있는지, 다시 말해 기계가 만들어 낸 것인지 여부를 판정하는 단계다. 두 번째 과제는 귀속(attribution)으로, 등록된 다수의 이용자 가운데 누가 해당 콘텐츠를 만들어 냈는지를 지목한다. 이때 후보가 되는 이용자는 미리 등록되어 있어야 하며, 판독하는 쪽은 그 명단을 알고 있는 상태라고 가정한다.

세 번째 과제는 추출(extraction)이며, 워터마크에 담긴 메시지를 읽어 내는 일에 해당한다. 앞의 귀속이 누구인지를 묻는다면, 추출은 워터마크에 무엇이 담겨 있는지를 묻는다. 네 번째 과제는 위치 지정(localization)으로, 콘텐츠의 어느 부분이 워터마크를 담고 있는지를 찾아낸다. 특히, 콘텐츠의 일부만 잘려 나간 상황에서도 그 위치를 특정할 수 있는지가 관건이 된다. 네 과제 중 위치 지정만 정보의 총량이 아니라 정보가 콘텐츠의 어디에 놓여 있는지를 다룬다.

이 네 과제는 합쳐 하나의 사다리를 이룬다고 할 수 있다. 아래에서 위로 올라갈수록 답을 얻는 데 더 많은 분량의 콘텐츠가 필요하기 때문이다. 맨 아래의 탐지는 가장 적은 분량으로 수행할 수 있으며, 워터마크를 어떤 방식으로 삽입했는지와도 무관하다. 반면, 그 위의 세 과제는 워터마크가 실제로 남겨 둔 정보에 따라 수행 가능 여부가 갈린다. 이 연구는 각 과제에 요구되는 콘텐츠 분량을 표본 복잡도(sample complexity)라는 척도로 계산해 낸다. 이는 목표한 정확도로 판정을 내리는 데 필요한 최소한의 자료 분량을 뜻한다.

[그림 1] 워터마크의 네 단계 포렌식 사다리와 단계별 요구 분량



출처: Xiaoyu Li 외 5인, "Watermark Forensics for Generative Models: An Information Theoretic Perspective", arXiv, 2026.07.14., <https://arxiv.org/pdf/2607.13003>

• 정보 프로파일의 개념과 역할

네 과제의 요구 분량을 계산하려면 먼저 공통 척도가 필요하다. 이 연구에 도입된 공통 척도는 정보 프로파일(information profile)이다. 워터마크가 삽입된 콘텐츠를 앞에서부터 한 조각씩 살펴본다고 할 때, 각 조각은 워터마크가 담고 있는 비밀 정보에 관해 조금씩 단서를 내놓는다. 어떤 조각은 많은 단서를 담고 있고 어떤 조각은 거의 없을 수 있는데, 정보 프로파일은 이 단서의 양을 위치별로 기록한 목록에 해당한다. 여기서 비밀 정보란 이용자의 신원이나 미리 답아 둔 메시지처럼, 나중에 읽어 내고자 하는 대상을 가리킨다.

이 목록은 두 가지 측면에서 해석할 수 있다. 한 가지 측면은 총량인데, 총량은 모든 위치의 값을 더한 것으로서 해당 콘텐츠에서 끌어낼 수 있는 비밀 정보의 양 전체에 해당한다. 이 연구에서는 총량을 포렌식 복원 한도(forensic-recovery budget)라 부른다. 포렌식 복원 한도는 콘텐츠가 만들어지는 순간 확정되므로, 이후 아무리 정밀하게 분석하더라도 애초에 삽입한 것보다 많은 정보를 얻어 낼 수는 없다. 분석 기법을 아무리 개선해도 넘어서 수 없는 상한선이 생성 시점에 이미 정해진다는 의미다. 다른 한 측면은 분포인데, 분포란 그 정보가 어느 위치에 몰려 있고 어느 위치에는 없는지를 나타낸다.

네 과제는 이 두 측면 가운데 어느 쪽을 읽느냐에 따라 갈린다. 귀속과 추출은 총량에 의존하며, 신원이나 메시지를 복원하기에 충분한 양이 쌓여야 비로소 수행할 수 있다. 위치 지정은 총량이 아니라 분포를 읽는 과제로, 정보가 어느 구간에 놓여 있는지가 관건이 된다. 탐지만은 예외인데, 워터마크가 삽입된 콘텐츠가 원래의 콘텐츠와 얼마나 다르게 보이는지에 따라 결정될 뿐 정보의 양과는 무관하다. 워터마크의 성능은 삽입 기법 자체가 아니라 워터마크가 남긴 정보 프로파일에 의해 결정되는 셈이다.

• 탐지와 귀속 사이에 놓인 간격

이 연구의 핵심 논지는 워터마크의 존재를 알아내는 일과 그 주인을 지목하는 일이 전혀 다른 과제라는 데 있다. 탐지는 워터마크가 있는지 없는지를 가리는 양자택일의 판정이므로, 등록된 이용자가 몇 명이든 요구되는 콘텐츠 분량이 달라지지 않는다. 반면, 귀속은 여러 후보 가운데 한 명을 지목하는 과제이므로, 후보의 수가 늘어날수록 더 많은 정보가 필요하다. 후보를 좁혀 나가는 데 그만큼의 정보가 소모되기 때문이다. 결국, 탐지가 가능해지는 시점과 귀속이 가능해지는 시점 사이에는 간격이 생긴다.

이 간격은 실제로 쓰이는 워터마크에서 확인된다. 널리 알려진 그린리스트(green-list) 방식은 콘텐츠를 생성할 때마다 사용할 수 있는 단어를 두 무리로 나누고 한쪽으로 치우치게 만드는 기법으로, 나중에 그 편향을 조사해 워터마크의 존재를 판정한다. 이 방식을 언어 모델에 적용한 실험에 따르면 탐지에 필요한 분량은 이용자 수와 무관하고 일정하게 유지되었다. 반면, 귀속에 필요한 분량은 이용자 수가 늘어남에 따라 함께 증가했다. 등록된 이용자가 많아질수록 두 과제 사이의 간격이 벌어진다는 뜻이다.

같은 현상은 통제된 조건의 모의실험에서도 나타난다. 탐지 성공률이 거의 최고 수준에 이른 시점에서도 귀속 정확도는 낮은 수준에 머물렀다. 탐지가 완전히 포화된 뒤에도 상당한 분량이 더 쌓이고 나서야 귀속 정확도가 의미 있게 올라갔다. 이는 두 과제가 필요로 하는 분량의 차이가 특별한 조건에서만 나타나는 현상이 아님을 보여 준다. 이 연구는 이러한 차이가 특정 기법의 결합이 아니라 정보량의 제약에서 비롯된다고 설명한다.

이 간격이 존재하는 구간에서는 두 과제의 결과가 엇갈린다. 해당 콘텐츠가 기계로 생성되었다는 사실은 확인되지만, 누가 만들었는지는 어떤 판독 방법으로도 알아낼 수 없다. 워터마크를 뚜렷이 지니면서도 익명으로 남는 상태인 셈이다. 이는 판독 기술이 아직 발전하지 않아서가 아니라, 애초에 그 분량의 콘텐츠에 담긴 정보가 한 명을 지목하기에 부족하기 때문이다. 이 연구는 정보가 부족한 상태에서 지목을 시도할 경우 무고한 이용자를 지목할 위험이 따른다고 지적하며, 그 위험을 일정 수준 이하로 억제하는 판독 방식을 함께 제시한다.

• 남은 한계와 제도적 함의

이 연구가 제시한 결과에는 분명한 적용 범위가 있다. 귀속에 필요한 분량을 계산한 부분은 콘텐츠가 원본 그대로 확보된 경우를 전제로 한다. 유출된 사본이 편집이나 압축을 거치면 워터마크와 대조하는 작업 자체가 어긋나므로, 같은 분량으로 귀속이 가능하다는 보장은 사라지게 된다. 이 연구에서는 편집을 전디는 조건에서도 동일한 요구 분량이 유지되는지 문제는 미해결 과제로 남겨 두고, 이 미해결 과제가 연구에서 제시한 프레임워크(framework)가 제기하는 가장 첨예한 문제라고 언급하고 있다.

판독 방식과 관련한 한계도 함께 지적된다. 후보 이용자를 걸러 내는 통상적인 방식은 콘텐츠와 어긋나지 않는 이용자를 모두 남기고 그 목록이 한 명으로 줄어들기를 기다린다. 그러나 이 방식대로라면 실제로 필요한 분량보다 많은 분량이 필요하다. 후보 목록이 여전히 많은 상태에서도 진짜 이용자를 골라낼 수 있는데, 목록이 줄어들 때까지 기다려야 하므로 분량이 낭비될 수밖에 없는 것이다. 이 연구에서는 여러 언어 모델을 대상으로 이러한 낭비의 정도를 측정하고, 그 값이 모델에 따라 크게 달라지지 않음을 밝혀 냈다. 즉, 낭비 정도가 모델에 따라 정해지는 특성이 아니라 자연어 자체의 특성에서 비롯되는 현상이라는 의미다.

이러한 한계는 제도 설계와도 맞닿아 있다. 유럽연합 인공지능법(AI Act) 제50조가 기계 판독이 가능한 형식의 표시(marks in a machine-readable format)를 요구하고, 중국의 심층합성 규정(China's Deep Synthesis provisions)이 눈에 띄는 라벨(noticeable labels)을 요구하고 있음을, 또 미국 캘리포니아주 SB 942가 잠재 고지(latent disclosure)를 요구하고 있음을 이 연구는 언급하고 있다. 이들 규정은 모두 콘텐츠에서 꺼내 읽을 수 있는 정보가 존재해야 한다는 전제를 깔고 있다. 그런데 콘텐츠 전체에 열게 퍼지는 방식의 워터마크는, 눈에 덜 띄게 만들수록 귀속에 필요한 분량이 커지는 구조를 지니고 있다. 법령을 통해 특정 기능의 이행을 요구할 수는 있으나, 법령이 정보량의 제약 자체를 없앨 수는 없다는 것이 이 연구의 결론이다.

정리하면 이 연구는 워터마크가 무엇을 할 수 있고 무엇을 할 수 없는지의 경계를 정보량의 관점에서 그려낸다. 탐지는 적은 분량으로 가능하지만, 귀속을 위해서는 이용자 수에 비례하는 분량이 필요하며, 좁은 구간에만 존재하는 워터마크로는 잘려 나간 콘텐츠에서 위치를 정밀하게 지목하기 힘들다는 것이다. 다만, 이 결론은 이론적 한계에 관한 것이어서, 실제 환경에서의 구현 방식에 관해서는 별도의 논의가 필요하다. 이어지는 본론 2에서는 화면 재촬영이라는 구체적인 유출 상황을 놓고 귀속을 실제로 구현한 연구를 살펴본다.

II. 본론 2: 화면 재촬영 환경에서의 유출 지점 추적

• 화면 재촬영이라는 유출 경로

중국 후난대학교(Hunan University)와 중국과학기술대학교(University of Science and Technology of China)가 2026년 7월 공개한 연구는 화면 촬영(screen-shooting)을 통한 유출에 대응하는 워터마크 기법을 제안한다. 화면 촬영이란 화면에 표시되는 콘텐츠를 카메라로 되찍어 사본을 만드는 행위를 가리킨다. 디스플레이 기기와 고화질 카메라가 널리 보급된 결과, 이 방식은 콘텐츠를 무단으로 확보하는 손쉬운 수단이 되었다. 이 연구는 화면 촬영이 저작권 침해와 정보 유출의 위험을 함께 낳는다고 지적한다.

이러한 경로에 대응하려면 워터마크가 촬영 과정을 견뎌야 한다. 화면을 촬영한 이미지는 원본과 여러 면에서 달라지는데, 촬영 각도에 따라 형태가 기울어지고 조명에 따라 밝기가 변하며 화면의 화소 격자와 카메라의 화소 격자가 어긋나 물결 모양의 무늬가 생기기도 한다. 이 같은 복합 변형을 견디며 워터마크를 복원하는 기법은 이미 여러 차례 제시된 바 있다. 이 연구가 제안하는 방식도 이러한 연구 흐름에 따라, 변형을 학습 단계에서 미리 재현해 두고 그런 조건에서도 워터마크를 되읽을 수 있도록 한다.

이 연구가 주목하는 지점은 워터마크를 읽은 다음 단계다. 기존 기법들은 하나의 콘텐츠에 하나의 워터마크를 삽입한 뒤 여러 화면에 동일하게 배포하는 것을 전제한다. 저작권자가 누구인지를 증명하는 데는 이 방식으로 충분하겠지만, 이 방식만으로는 여러 화면 가운데 어느 화면에서 유출이 일어났는지는 알 수 없다. 이 연구는 다중 화면 환경의 워터마크가 동일한 저작권 정보를 유지하면서도 화면을 구별할 수 있어야 한다는 점을 지적한다.

이 연구가 제안하는 공모 저항형 이미지 무관 워터마크(Collusion-Resistant Image-Agnostic Watermarking, 이하 CoMSMark)는 이러한 요구에 대응하기 위한 기법이다. CoMSMark는 저작권 정보와 화면 식별자(screen ID)를 함께 처리해, 같은 저작권 정보를 유지하면서도 화면마다 서로 다른 형태의 워터마크를 생성한다. 유출된 사본에서 저작권 정보를 읽어 권리자를 확인하고, 동시에 화면 식별자를 읽어 유출 지점을 특정하는 구조다. 다만, 화면마다 다른 워터마크를 배포하는 방식은 그 자체로 새로운 위험성을 초래하는데, 이 연구의 상당 부분은 그 위험성에 대응하는 데 할애하고 있다.

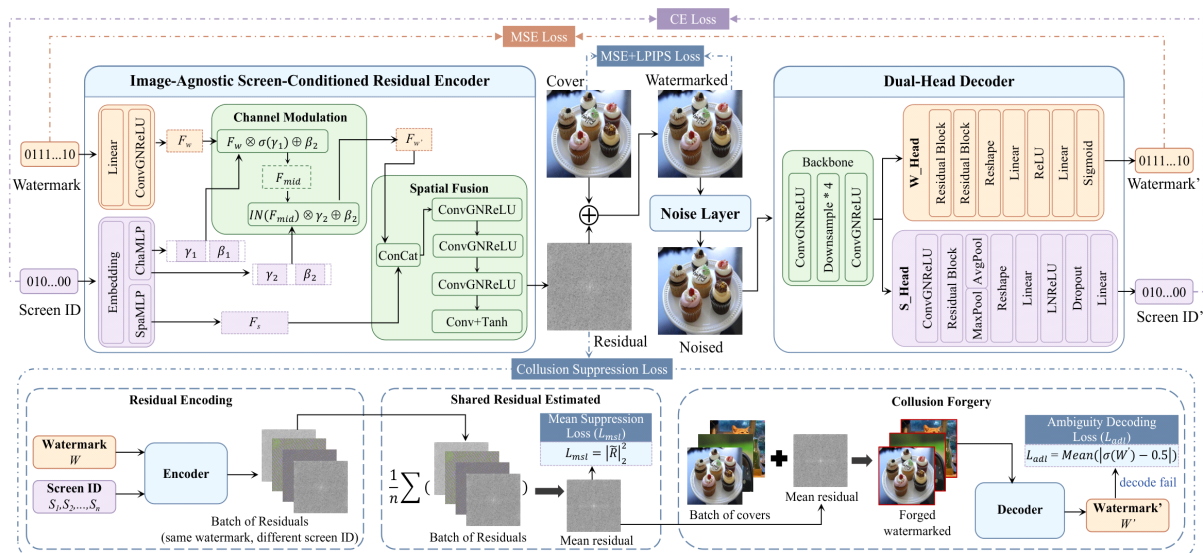
• 다중 화면 공모 공격과 CoMSMark의 대응 설계

화면마다 다른 워터마크를 배포하면 유출 지점을 특정할 수 있게 되지만, 동시에 공격자에게 여러 사본을 비교할 기회가 생긴다. 여러 배포처에 흩어져 있던 사본이 한 공격자에게 모이면, 각 사본에서 워터마크가 없는 이미지를 빼고 그렇게 얻은 값들을 평균하는 방식으로 워터마크의 형태를 추정할 수 있다. 사본마다 다른 요소는 서로 상쇄되어 없어지는 반면, 모든 사본에 공통으로 들어 있는 부분은 그대로 남기 때문이다. 원본을 확보하지 못한 상태에서도 사본을 충분히 모으기만 하면 워터마크 추정이 가능하다는 뜻이다. 이 연구는 이러한 위협을 다중 화면 공모 공격(multi-screen collusion attack)이라 부르며, 화면 촬영 워터마크 연구에서 이 문제가 명시적으로 다뤄진 적이 없다고 지적한다.

이 공격은 두 가지 상황으로 나뉜다. 하나는 사본마다 담긴 이미지가 서로 다른 경우로, 이미지의 내용이 사본별로 달라 평균을 낼 때 함께 상쇄된다. 다른 하나는 같은 이미지가 여러 화면에서 촬영된 경우로, 이미지의 구조가 사본 사이에서 일치하기 때문에 상쇄되지 않고 추정값에 남는다. 후자에서는 공격자가 추정한 형태를 얼마나 강하게 적용할지 조절해, 남아 있는 이미지 내용의 영향을 통제할 수 있다.

추정된 형태는 두 가지 방향으로 악용될 수 있다. 하나는 제거로서, 워터마크가 삽입된 이미지에서 추정한 형태를 덜어내 워터마크를 읽을 수 없게 만드는 과정이다. 다른 하나는 위조로서, 워터마크가 없는 이미지에 추정한 형태를 더해 정당한 배포본처럼 꾸미는 과정이다. 이 연구는 위조가 거짓된 소유권 주장으로 이어질 수 있다는 점도 함께 지적하고 있다.

[그림 2] CoMSMark 프레임워크의 구조



출처: Mingyue Chen 외 4인, "Collusion-Resistant Image-Agnostic Watermarking for Multi-Screen Shooting", arXiv, 2026.07.31., <https://arxiv.org/pdf/2607.23553>

CoMSMark는 이 공격의 전제를 무너뜨리는 방식으로 대응한다. 공모 공격이 성립하는 이유는 여러 사본의 워터마크가 통계적으로 서로 닮아 있어 평균을 냈을 때 공통 부분이 남기 때문이다. 이에 화면 식별자를 조건으로 받아들이며 화면마다 형태가 다른 워터마크를 만들어 내되, 담긴 저작권 정보는 동일하게 유지한다. 나아가 학습 과정에서 여러 화면의 워터마크를 평균한 값이 작아지도록 유도해 공통 부분을 줄인다. 같은 단계에서 위조된 사본을 직접 만들어 보고, 그 사본에서는 워터마크를 판독하지 못하도록 함께 학습시킨다.

읽어내는 쪽에도 별도의 설계가 적용된다. 유출된 사본에서 저작권 정보와 화면 식별자를 각각 담당하는 두 갈래로 나누어 판독하는 구조를 두어, 권리자 확인과 유출 지점 특징을 함께 수행한다. 아울러 워터마크를 이미지 내용과 무관하게 생성하도록 설계해, 이미지마다 새로 계산하지 않고 만들어 둔 워터마크를 여러 이미지에 그대로 적용할 수 있게 한다. 대규모 배포 환경에서 처리 부담을 줄이기 위한 선택이다.

• 검증 결과와 한계

CoMSMark의 성능은 실제 기기를 사용한 실험으로 확인되었다. 노트북 화면에 이미지를 표시하고 삼각대에 고정된 스마트폰으로 촬영하는 방식이며, 촬영 거리와 각도를 달리한 조건에서도 워터마크와 화면 식별자를 모두 높은 정확도로 읽어냈다. 기기 조합을 바꾸어도 실험 결과는 안정적으로 유지되었다. 화면마다 다른 워터마크를 만들어 넣더라도 저작권 정보의 판독이 저해되지 않고, 반대로 저작권 정보가 달라져도 화면 식별에 영향을 주지 않는 것으로 나타났다. 두 정보를 나누어 읽는 구조가 서로를 방해하지 않고 작동한다는 의미다.

공모 공격에 대한 방어력도 함께 검증되었다. 여러 사본을 모아 워터마크를 추정해낸 뒤 이를 털어내는 제거 공격에서, 사본마다 이미지가 다른 상황에서는 일부 기존 기법의 판독 정확도가 무작위로 찍는 수준에 근접한 반면 CoMSMark는 높은 정확도를 유지했다. 위조 공격에서도 기존 기법들은 워터마크가 없는 이미지가 정당한 배포본으로 판독되는 비율이 높아졌으나, CoMSMark에서는 그 비율이 무작위 수준에 머물렀다. 워터마크가 삽입된 이미지의 화질 역시 기존 기법과 견줄 만한 수준으로 유지되었다. 아울러 공격 강도를 높일수록 공격 대상 이미지의 화질이 크게 나빠지는 현상이 확인되어, 공격의 성공과 이미지의 사용 가치가 상충 관계에 있음이 드러났다.

이 기법의 적용 범위는 실험 조건에 나타난다. 검증은 정지된 이미지를 대상으로 수행되었으며, 워터마크는 정해진 길이의 값으로 두고 화면 식별자는 미리 정한 개수 안에서 부여했다. 아울러 화면 식별자를 읽어 유출 지점을 특정하는 구조이므로, 어느 배포처에 어떤 식별자를 부여했는지 확인될 수 있어야 작동한다. 이 연구는 이러한 조건 아래에서 기존 기법보다 공모 공격에 강하면서도 화질과 판독 성능을 함께 유지했다고 정리한다.

Ⅲ. 결론 및 전망

• 두 연구가 공통으로 드러낸 출처 추적의 조건

두 연구는 접근 방식이 다르지만 같은 결론으로 향하고 있다. 첫 번째 연구는 워터마크가 존재한다는 사실을 확인하는 일과 그 출처를 지목하는 일이 서로 다른 조건 위에서 작동한다는 점을 정보량의 관점에서 밝혔다. 두 번째 연구는 화면 촬영이라는 구체적 유출 경로를 놓고, 화면 식별자를 함께 삽입해 출처를 지목하는 방식을 실제로 구현했다. 첫 번째 연구는 무엇이 가능하고 무엇이 불가능한지의 경계를 한정했고, 두 번째 연구는 그 경계 안에서 작동하는 수단을 제시한 셈이다.

두 연구가 공통적으로 보여 주는 것은 출처 추적이 저절로 따라오지 않는다는 사실이다. 콘텐츠에 워터마크를 삽입해 두었다고 해서 유출자가 자동으로 드러나지는 않으며, 이용자를 구별할 정보가 충분히 담겨 있어야 비로소 지목이 가능해진다. 첫 번째 연구는 그 정보가 부족한 구간에서 무리한 지목이 무고한 이용자를 겨냥할 위험을 지적했다. 두 번째 연구는 이용자를 구별하려는 시도가 오히려 여러 사본을 비교할 기회를 공격자에게 제공한다는 점을 드러냈다. 출처를 추적하려는 설계가 그 자체로 새로운 위험을 만들어 낸다는 문제의식이 양쪽에 공통으로 깔려 있다.

'오디세이' 사안에 비추어 보면 두 연구의 의미가 분명해진다. 배급사가 취한 조치는 이미 확산된 복제본을 지우는 데 초점이 있었고, 그 사본이 어디에서 비롯되었는지에 관해서는 어떤 단서도 남기지 않았다. 유통된 사본 자체가 출처에 관한 정보를 지니고 있었다면, 삭제와는 다른 별개의 대응 경로가 열린다. 다만, 이러한 별개 대응 경로가 아무런 조건 없이 열리지는 않는다는 점도 두 연구는 함께 보여 준다. 편집이나 압축을 거친 사본에서 출처를 특정하는 문제는 아직 해결되지 않았고, 화면 식별자를 활용하는 기법도 배포처가 사전에 등록되어 관리되는 환경을 전제로 하고 있다.

• 남은 기술적 과제와 논의의 쟁점

기술 측면에서 남은 과제는 변형을 거친 사본에서의 출처 지목이다. 첫 번째 연구는 사본이 원본 그대로 확보된 경우를 전제해 요구 분량을 계산했고, 편집을 견디는 조건에서도 같은 결과가 유지되는지를 미해결 과제로 남겨 두었다. 그런데 실제 유출은 대부분 재압축이나 부분 절취를 거친 사본에서 발견된다. 두 번째 연구가 촬영 과정의 변형을 학습 단계에서 미리 재현해 대응한 방식은, 예상되는 변형을 설계에 반영한 한 가지 사례라고 할 수 있다.

배포 관리 체계와 관련한 조건도 확인할 수 있다. 화면 식별자를 활용하는 방식은 어느 배포처에 어떤 식별자를 부여했는지가 기록으로 남아 있어야 제대로 작동한다. 기록이 없으면 유출된 사본에서 식별자를 읽어내더라도 그것이 누구의 것인지 특정할 수 없다. 워터마크 기술을 적용하려면 삽입 기법의 선택만으로는 부족하고, 배포처 등록과 식별자 관리 절차를 함께 진행해야 한다.

첫 번째 연구는 제도와 관련된 쟁점도 하나 제기하고 있다. 첫 번째 연구는 여러 국가의 규정이 콘텐츠에서 꺼내 읽을 수 있는 정보의 존재를 전제하고 있다고 환기시키면서도, 그러한 요구가 정보량의 제약 자체를 없애지는 못한다고 결론짓고 있다. 워터마크의 존재 확인과 출처 지목이 서로 다른 조건 위에서 작동한다는 사실은, 워터마크에 어느 수준의 기능을 기대할 수 있는지를 둘러싼 논의로 이어질 수 있다. 첫 번째 연구에서는 이러한 논의의 결론을 법률 해석이 아니라 기술적 귀결로 제시하고 있다.

• 유출 대응 체계에서 워터마크가 지닌 위치와 향후 과제

두 연구는 워터마크 논의의 무게중심이 옮겨 가고 있음을 보여 준다. 구체적으로 말해서, 워터마크가 있는지를 가려내는 단계에서 그 워터마크가 누구의 것인지를 지목하는 단계로 관심이 이동하고 있다는 것이다. 첫 번째 연구가 그 이동에 따르는 대가를 정보량으로 계산해 냈고, 두 번째 연구는 화면 촬영이라는 특정 경로에서 그 이동을 실제로 구현했다. 다만, 두 연구 모두 출처 지목이 조건 없이 가능해지는 것은 아니라는 점을 말하고 있다.

이러한 흐름 속에서 워터마크는 삭제 중심 대응을 대체하는 수단이 아니라 그것을 보완하는 수단으로 자리한다. 이미 확산된 침해물을 정리하는 일과 그 침해물이 어디에서 비롯되었는지를 밝히는 일은 서로 다른 과제이며, 워터마크는 후자에 필요한 근거를 콘텐츠 자체에 남겨 둔다. 다만, 그 근거가 실제로 읽히려면 확보된 사본의 분량과 상태, 배포처 관리 체계라는 조건이 함께 충족되어야 한다. 앞으로의 연구가 향하는 지점도 이러한 조건을 완화하는 데 놓여 있으며, 편집에 강건한 출처 특정과 대규모 배포 환경에서의 식별자 운용이 이러한 조건 완화의 핵심이라고 하겠다.

참고문헌

- Carly Thomas, “Universal Responds to ‘The Odyssey’ Leaks on Social Media”, The Hollywood Reporter, 2026.07.26., <https://www.hollywoodreporter.com/movies/movie-news/universal-responds-the-odyssey-leaks-x-social-media-1236656766/>
- Washington Times, “Bootleg copies of Nolan’s ‘The Odyssey’ flood X despite takedowns”, 2026.07.28., <https://www.washingtontimes.com/news/2026/jul/28/bootleg-copies-nolans-odyssey-flood-x-despite-takedowns/>
- Xiaoyu Li 외 5인, “Watermark Forensics for Generative Models: An Information-Theoretic Perspective”, arXiv, 2026.07.14., <https://arxiv.org/pdf/2607.13003>
- Mingyue Chen 외 4인, “Collusion-Resistant Image-Agnostic Watermarking for Multi-Screen Shooting”, arXiv, 2026.07.31., <https://arxiv.org/pdf/2607.23553>