일반논문 (Regular Paper)
방송공학회논문지 제31권 제2호, 2026년 3월 (JBE Vol.31, No.2, March 2026)
https://doi.org/10.5909/JBE.2026.31.2.318
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

미디어 콘텐츠 보안을 위한 양방향 객체 추적 및 최적 식별 모델 조합 기반의 강건한 동적 워터마킹 연구

강 자 원[a)], 조 동 헌[a)], 최 승 관[a)‡]

A Study on Robust Dynamic Watermarking for Media Content via Bidirectional Object Tracking and Optimal Identification Model Integration

Jawon Kang[a)], Dongheon Jo[a)], and Seungkwan Choi[a)‡]

요 약

본 논문은 동적 미디어 환경에서 객체의 움직임과 가려짐에 취약한 기존 워터마킹의 한계를 극복하기 위해, 양방향 객체 추적 알고리즘 기반의 강건한 동적 워터마킹 파이프라인을 제안한다. 제안 시스템은 YOLOv8과 CLIP을 결합하여, 범용 Zero-shot 모델인 OWL-ViT2 대비 비디오 보안 도메인에서 우수한 식별 강건성을 확보하였다. 특히 SAM2를 활용한 '참조 프레임 기반 양방향 시간적 전파' 로직을 설계하여 복잡한 장면에서도 워터마크가 적용된 마스크를 정밀하게 유지한다. 또한 해밍 거리 기반 검증 체계를 도입하여 영상 변조 시 발생하는 비트 오류에 대한 기술적 엄밀성을 강화하였다. OTB 벤치마크 실험 결과, 제안 기법은 인물 추적 IoU에서 Baseline 대비 약 2.6배 향상된 성능을 보였으며, 각종 변조 공격에서도 96% 이상의 높은 검출률을 기록하였다. 본 연구는 최신 비전 모델의 최적 조합을 통해 실시간 방송 보안 환경에 즉시 적용 가능한 기술적 토대를 마련하였다.

Abstract

This paper proposes a robust dynamic watermarking pipeline based on a bidirectional object tracking algorithm to overcome the limitations of conventional watermarking in dynamic media environments. By integrating YOLOv8 and CLIP, the system achieves superior identification robustness in the video security domain compared to general-purpose zero-shot models like OWL-ViT2. Specifically, we designed a 'Reference-based Bidirectional Temporal Propagation' logic using SAM2, ensuring precise maintenance of watermark masks even amidst complex movements and occlusions. Furthermore, technical rigor was enhanced through a verification framework based on Hamming distance to mitigate bit errors caused by video manipulation. Experimental results using the OTB benchmark demonstrated that the proposed method achieved a 2.6-fold increase in tracking IoU compared to the baseline and maintained a detection rate exceeding 96% under various distortion attacks. This study provides a practical technical foundation for real-time broadcasting security by identifying the optimal combination of state-of-the-art vision models.

Keyword : Bidirectional Object Tracking, Dynamic Watermarking, Robustness Analysis, Spatio-temporal Consistency, Deepfake Defense

Ⅰ. 서 론

디지털 전환 이후 VOD, OTT, 소셜 미디어를 포함한 다양한 미디어 콘텐츠 유통이 활성화되면서 전례 없는 부가가치가 창출되고 있다. 그러나 이러한 다매체 환경은 불법 복제, 악의적인 딥페이크 제작, 무단 편집 등 저작권 침해 및 콘텐츠 변조의 위험성을 기하급수적으로 증가시켰다[1]. 기존의 정적 워터마킹 기술은 프레임 전체에 일괄적으로 정보를 삽입하기 때문에 공격자가 삽입 영역을 예측하여 쉽게 제거할 수 있으며, 특히 객체의 움직임이 빈번한 동적 콘텐츠 환경에서는 시각적 왜곡과 보안 정보 유실이라는 상충하는 한계를 지닌다.

최근 이를 해결하기 위해 객체 인지 기반의 동적 워터마킹 연구가 진행되고 있으나, 기존 방식은 객체의 급격한 움직임이나 가려짐(Occlusion) 발생 시 시간적 일관성(Temporal Consistency)을 잃어 워터마크 추출 성능이 급격히 저하되는 난제를 안고 있다. 또한, 범용적인 Zero-shot 탐지 모델은 비디오 보안이라는 특수 도메인에서 실시간성과 식별 강건성을 동시에 확보하는 데 한계를 보인다.

본 논문에서는 이러한 기술적 난제를 해결하기 위해 ‘양방향 객체 추적 및 최적 식별 모델 조합 기반의 강건한 동적 워터마킹 파이프라인’을 제안한다. 본 연구에서 제안하는 파이프라인은 단순히 개별 AI 모델을 나열한 워크플로우를 넘어, 비디오 보안 환경에 최적화된 모델 스택(YOLOv8 & CLIP)을 규명하고, 시공간적 일관성을 극대화하기 위한 ‘참조 프레임 기반 양방향 시간적 전파(Bidirectional Temporal Propagation)’ 알고리즘을 핵심 동력으로 삼는다. 이는 콘텐츠 내 핵심 객체를 정밀하게 인지하고 추적하여 보안 정보를 동적으로 매핑함으로써, 변조 공격에 능동적으로 대응하는 보안 체계를 지향한다.

본 논문의 주요 기여점 및 차별성은 다음과 같다.

첫째, 비디오 보안 도메인에 최적화된 식별 모델 조합의 검증 및 적용이다. 최신 Zero-shot 탐지 모델인 OWL-ViT2 등과의 비교 분석을 통해, 비디오 환경에서 식별 강건성을 극대화할 수 있는 YOLOv8과 CLIP의 조합적 우수성을 실험적으로 입증하였다. 이를 통해 단순한 시스템 통합을 넘어 보안 실무 환경에 최적화된 기술적 판단 근거를 제시하였다.

둘째, SAM2 기반의 양방향 시간적 전파 로직 설계를 통한 추적 강건성(Robustness) 확보이다. 단방향 추적의 고질적 문제인 객체 유실 및 추적 실패를 극복하기 위해, 핵심 프레임을 기점으로 순방향과 역방향을 동시에 추적하는 알고리즘을 제안하였다. 이를 통해 크롭(Crop)이나 객체 가림과 같은 회피 공격 상황에서도 워터마크가 삽입된 픽셀 영역을 정밀하게 유지하는 높은 시공간적 일관성을 달성하였다.

셋째, 학술적 벤치마크를 통한 정량적 검증 및 실무적 신뢰성 확보이다. 특정 데이터에 국한되지 않고 공인된 OTB(Object Tracking Benchmark) 데이터셋을 활용하여 제안 알고리즘의 우수성을 객관적으로 입증하였다. 또한, 해밍 거리(Hamming Distance) 기반의 이진 정보 복원 알고리즘을 적용하여 영상 훼손 시 발생하는 비트 오류에 대한 기술적 엄밀성을 강화하였다.

결과적으로 본 연구는 콘텐츠 인지형 선택적 보호 기술을 통해 지능화되는 저작권 침해에 대응하고, 실시간성 및 강건성을 동시에 확보한 실효성 있는 미디어 보안 체계의 새로운 기술적 방향을 제시한다.

a) 서강대학교 가상융합전문대학원 가상융합테크놀로지(Virtual Convergence Technology, Sogang University)
‡ Corresponding Author : 최승관(Seungkwan Choi)
E-mail: csk0123@sogang.ac.kr
Tel: +82-2-705-8848
ORCID: https://orcid.org/0009-0003-9620-2079
※본 연구는 과학기술정보통신부 및 정보통신기획평가원의 ‘메타버스 융합대학원’ 과제(RS-2022-00156318)와, 문화체육관광부 및 한국콘텐츠진흥원의 ‘메타버스에서 저작권 보호 및 이용 활성화 기술 개발’ 과제(RS-2023-00219237)의 지원을 받아 수행되었다.
· Manuscript February 19, 2026; Revised March 10, 2026; Accepted March 10, 2026.

Ⅱ. 관련 연구 및 배경 기술

제안하는 지능형 동적 워터마킹 파이프라인은 영상 내 특정 객체를 지속적으로 추적하는 기술과 식별 정보를 콘텐츠에 삽입하는 기술의 유기적 결합을 기반으로 한다. 본 장에서는 참고문헌을 바탕으로 기존 연구의 기술적 한계를 분석하고, 이를 극복하기 위해 본 연구에서 채택한 핵심 방법론의 타당성을 고찰한다.

1. 디지털 워터마킹 기술의 발전과 한계

1.1 고전적 주파수 영역 및 객체 기반 방식[2][3]

초기 연구는 DCT나 DWT 등 주파수 계수를 변형하여 강인성을 확보하는 데 집중했으나, 영상의 시각적 문맥을 고려하지 못해 특정 객체(인물, PPL 등)를 선택적으로 보호하는 데 한계를 보였다. 이후 SVD를 활용한 관심 영역(ROI) 삽입 방식이나 특징점 매칭 기반 연구들이 제안되었으나[4], 비디오 환경에서 객체의 급격한 형태 변화나 가려짐(Occlusion)이 발생할 경우 추적 일관성을 유지하지 못해 보안 정보가 유실되는 고질적인 문제를 안고 있다.

1.2 딥러닝 기반 동적 워터마킹(WAM)

최근 제안된 WAM(Watermark Anything Model)은 객체 영역에 동적으로 정보를 삽입하며[5] 움직임과 가려짐에 대한 강인성을 비약적으로 향상시켰다. 그러나 WAM은 단일 프레임 내의 마스크 기반 삽입에 최적화되어 있어, 장시간 비디오 스트림에서 동일 객체에 대한 식별 정보를 지속적으로 유지하기 위해서는 외부의 정밀한 추적 및 식별 로직과의 결합이 필수적이다.

2. 비디오 보안을 위한 탐지 및 식별 기술의 최적화

2.1 오픈 어휘 탐지(OWL-ViT2)와 YOLO+CLIP 조합 비교

최신 OWL-ViT2와 같은 Zero-shot 탐지 모델은 범용적인 객체 인식에 강점이 있으나, 비디오 보안 실무에서 요구되는 실시간 처리 속도와 특정 타겟에 대한 식별 안정성 측면에서는 한계를 보인다. 본 연구는 고속 stage 1 탐지기인 YOLOv8[6]로 후보 영역을 선별하고, CLIP[7]의 시각-언어 임베딩을 통해 정밀 매칭을 수행하는 조합을 채택하였다. 이는 범용 모델 대비 비디오 보안 도메인에서 보다 높은 식별 강건성을 제공함을 실험적으로 입증하였다.

2.2 InsightFace를 활용한 정밀 안면 식별

안면 보안이 중요한 미디어 환경을 고려하여 InsightFace[8]를 병행 채택하였다. ArcFace 손실 함수 기반의 특징 추출 방식은 클래스 간 구별력을 극대화하여, 딥페이크와 같은 정교한 안면 변조[9] 상황에서도 원본 객체를 정확히 특정할 수 있는 학술적 근거를 제공한다.

3. 시공간적 일관성 확보를 위한 추적 기술(SAM2)

비디오 객체 추적의 최신 기준점인 SAM2[10]는 제로샷 세그멘테이션 능력을 통해 복잡한 배경에서도 정밀한 마스크를 생성한다. 그러나 단방향 전파(Forward Propagation) 방식은 초기 탐지 실패 시 이후 모든 프레임에서 추적이 유실되는 '오류 누적(Error Drift)' 문제가 발생한다. 본 연구에서는 이를 해결하기 위해 참조 프레임을 기점으로 순방향과 역방향을 동시에 추적하는 '양방향 시간적 전파(Bidirectional Temporal Propagation)' 로직을 설계하였다. 이는 기존 SAM2[11]의 성능을 보안 도메인에 맞게 고도화한 것으로, 미디어 콘텐츠의 시공간적 무결성을 보장하는 핵심 기술적 차별점이다.

Ⅲ. 제안하는 지능형 동적 워터마킹 시스템

본 연구에서 제안하는 시스템은 개별 AI 모델의 단순 통합을 넘어, 미디어 보안 환경에서의 강건성을 극대화하기 위해 설계된 엔드투엔드(End-to-End) 파이프라인이다. 특히, 특정 객체에 대한 식별 일관성을 유지하기 위한 '양방향 객체 추적 로직'과 데이터 손실에 대응하는 'DB 연동형 검증 체계'를 핵심 차별점으로 가진다.

1. 객체 추적 기반 비디오 워터마킹 핵심기술

본 연구에서 제안하는 전체 시스템 아키텍처는 그림 1과 같다. 제안 파이프라인은 원본 영상으로부터 보호 대상 객체를 탐지 및 식별하고(Step 3-4), 식별된 고유 ID에 대응하는 보안 페이로드를 생성하여(Step 1-2), 시공간적으로 일관된 영역에 워터마크를 삽입 및 검증하는(Step 5-6) 유기적 엔드투엔드(End-to-End) 구조를 가진다. 본 시스템은 데

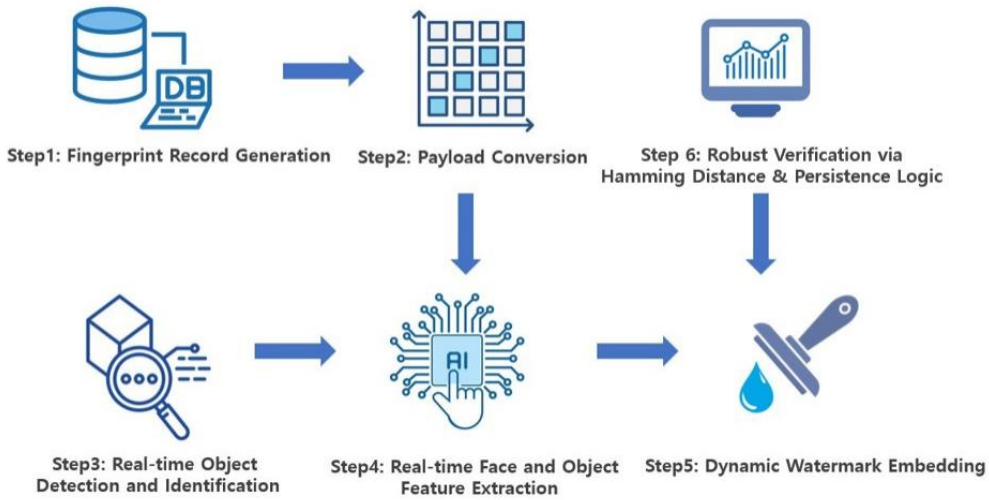


그림 1. DB 연동형 워터마크 삽입 워크플로우
Fig. 1. Database-Linked Watermark Embedding Workflow

이터 흐름에 따른 6개의 세부 절차(Step 1~6)로 구성되며, 이는 기술적 역할에 따라 핵심 AI 모델 기반의 4개 단계(Stage 1~4)로 분류되어 상호작용한다. 특히, 각 단계는 후술할 양방향 추적 로직과 지속성 삽입 알고리즘을 통해 동적 미디어 환경에서의 강건성을 극대화하도록 설계되었다.

그림 1에 제시된 지능형 동적 워터마킹 파이프라인의 세부 단계별 구동 원리는 다음과 같다.

1) Step 1(핑거프린트 레코드 생성): 워터마크 삽입 작업 시작 시 사용자 정보 및 원본 소스를 SQLite DB에 기록하고, 주기적으로 갱신되는 고유한 fingerprint_id를 생성한다.
2) Step 2(페이로드 변환): 생성된 ID를 32비트 고정 길이 이진 문자열로 변환하여 실질적인 워터마크 페이로드로 정의한다.
3) Step 3(실시간 객체 탐지 및 추적): SAM2 기반의 객체 분할과 InsightFace/CLIP을 이용한 특징 추출을 병행한다. 특히 본 연구에서 제안하는 양방향 추적(Reverse-forward tracking) 로직을 통해 가려짐(Occlusion) 상황에서도 객체를 안정적으로 추적한다.
4) Step 4(동적 워터마크 삽입): WAM 모델을 통해 정밀 마스크 영역에 페이로드를 삽입한다. 이때 지속성 삽입(Embedding Persistence) 로직을 적용하여 분할 마스크의 미세한 깜빡임(Flickering) 현상에 따른 보안 공백을 차단한다.

1.1 단계별 AI 모델 통합 및 데이터 전파 로직

제안 시스템은 기술적 역할에 따른 각 단계(Stage)가 유기적으로 데이터를 전파하는 선순환 구조로 설계되었다. Stage 1(YOLOv8, Step 3의 객체 탐지)에서 추출된 객체의 바운딩 박스는 Stage 3(SAM2, 시공간적 추적)의 정밀 추적을 위한 시각적 프롬프트(Visual Prompt)로 변환되어 전달되며, Stage 2(CLIP, Step 4의 정밀 식별)의 결과는 DB 내 고유 ID와 매칭되어 Stage 4(WAM, Step 5의 동적 삽입)의 보안 페이로드 생성을 위한 근거가 된다.

1.2 최적 식별 모델 조합: YOLOv8 & CLIP

비디오 보안 도메인에서 실시간성과 강건성을 동시에 확보하기 위해 YOLOv8과 CLIP의 조합을 채택하였다. 범용 Zero-shot 탐지 모델과 달리, 본 조합은 고속 탐지(YOLOv8) 후 참조 이미지 집합 R={r_{1}, r_{2}, ..., r_{n}}과의 코사인 유사도(s_{i}) 매칭을 통해 특정 보호 대상을 정밀하게 식별한다.

$$s_i = \max_j \frac{c_i \cdot r_j}{||c_i|| \, ||r_j||} \tag{1}$$

1.3 양방향 시간적 전파(Bidirectional Temporal Propagation) 알고리즘

단방향 추적 시 발생하는 오류 누적과 객체 가림 문제를 해결하기 위해 양방향 추적 기법을 제안한다. 영상 내에서 신뢰도가 가장 높은 최적의 키프레임을 선정하여 SAM2를 초기화한 후, 이를 기점으로 순방향(Forward)과 역방향(Reverse)으로 마스크를 전파한다.

$$M_{t\pm1} = SAM2_{propagate}(M_t, I_{t\pm1}) \quad (2)$$

이러한 양방향 전파 방식은 객체가 영상 중간에 등장하거나 일시적으로 사라지는 복잡한 방송 장면에서도 워터마크 영역의 일관성을 보장한다.

1.4 동적 워터마크 삽입 로직

제안 시스템의 마지막 단계(Stage 4)인 워터마크 삽입 프로세스는 SAM2로 추적된 정밀 객체 영역에 WAM (Watermark Anything Model)[12]을 사용하여 워터마크를 동적으로 삽입한다. 이때 배경 왜곡을 최소화하고 연산 효율을 높이기 위해 마스크 기반 국소 삽입(Mask-based Local Embedding) 방식을 적용한다. 추적된 객체의 바운딩 박스 중심을 기준으로 512 x 512 픽셀의 정사각형 윈도우를 설정하고, 해당 영역 내에서만 WAM의 출력과 원본 프레임을 합성하여 배경 왜곡을 방지한다.

$$I_w = M \odot WAM(I,b) + (1-M) \odot I \quad (3)$$

여기서 I는 원본프레임, b는 이진 페이로드, M은 삽입 마스크를 의미하며, $\odot$ 은 요소별 곱(Hadamard product)을 뜻한다. 또한, 삽입되는 워터마크 메시지는 핑거프린트 스케줄링 정책에 따라 일정 시간 간격(예: 5초)마다 갱신되며, 각 ID는 32비트 이진 페이로드로 변환되어 영상의 특정 구간 유출 경로를 세밀하게 추적한다.

2. DB 연동형 콘텐츠 보호 및 검증 파이프라인

2.1 지속성 삽입(Embedding Persistence) 로직

그림 1의 Step 5(워터마크 삽입) 및 Step 6 과정에서 객체 식별 후 WAM을 통해 워터마크를 삽입할 때, 분할 마스크의 미세한 깜빡임(Flickering)으로 인한 보안 공백을 방지하기 위해 지속성 삽입 로직을 적용한다. 이는 추적 알고리즘의 일시적 오류로 마스크가 끊기더라도 일정 프레임 동안 동일 위치에 워터마크를 지속하여 공격자의 회피 시도를 무력화한다.

2.2 해밍 거리 기반 오류 정정 및 검증

그림 1의 Step 6(최종 검증) 단계에서 추출된 이진 페이로드는 영상 압축이나 노이즈 공격으로 인해 비트 오류를 포함할 수 있다. 본 시스템은 이를 극복하기 위해 DB에 저장된 fingerprint_id와 추출된 페이로드 간의 해밍 거리(Hamming Distance)[13]를 계산하여, 허용 오차(Tolerance) 이내의 가장 유사한 ID를 유효한 정보로 확정함으로써 검증의 신뢰성을 높인다.

2.3 워터마크 검출 및 검증 프로세스

불법 유통이 의심되는 영상에서 워터마크를 검출하고 정보를 확인하는 과정은 삽입의 역순으로 진행되며, 오류 보정 메커니즘을 포함한다.

1) 목표 객체 재식별: 검증할 영상에서도 삽입 과정과 동일하게 SAM2를 이용해 객체 영역을 분할한 뒤, InsightFace/CLIP을 사용하여 워터마크가 삽입되었을 것으로 추정되는 목표 객체를 찾는다.
2) 워터마크 페이로드 추출: 목표 객체가 식별된 위치에서 WAM의 검출 기능을 이용해 이진 페이로드를 추출한다. 영상이 압축 등의 공격을 받은 경우, 이 페이로드는 일부 비트 오류를 포함할 수 있다.
3) 해밍 거리 기반 ID 매칭: 추출된 이진 페이로드를 데이터베이스에 저장된 모든 유효 fingerprint_id의 이진 문자열과 비교하여 해밍 거리(두 이진 문자열 간 다른 비트의 수)를 계산한다.
4) 정보 조회 및 검증: 계산된 해밍 거리가 가장 작고, 사전에 설정된 허용 오차(tolerance) 이내일 경우, 해당 ID를 유효한 핑거프린트로 확정하고 데이터베이스에서 일치하는 사용자 정보 및 원본 텍스트를 조회한다.

3. 시스템 구현 및 최적화 전략

3.1 시스템 개발 환경

제안하는 '동적 콘텐츠 보호 파이프라인'의 프로토타입은 이론적 설계를 넘어 실제 방송 및 VOD 유통 환경에서의 실효성(Practicality)과 동작 가능성(Viability)을 검증하기 위해 다음과 같은 환경에서 구현되었다. 전체 시스템은 다중 AI 모델 통합에 따른 연산 병목 현상을 해결하고, 고해상도 영상 처리 시의 안정적인 데이터 파이프라이닝에 중점을 두어 설계되었다. 구체적인 시스템 사양 및 특징은 다음과 같다.

표 1. 시스템 개발 환경의 HW 구성 및 SW 스펙
Table 1. Hardware Configuration and Software Specifications

Category	Item	Specifications
HW	GPU	NVIDIA GeForce RTX 4060 (VRAM 8GB)
	CPU	Intel Core i5-13700K
	RAM	32GB DDR5
SW	Language	Python (3.13)
	AI Framework	PyTorch (1.13.1)
	Core Model	YOLOv8, InsightFace, CLIP, WAM
	Data Processing	OpenCV, FFmpeg
	Database	SQLite3

1) 하드웨어 최적화: NVIDIA GeForce RTX 4060 GPU와 8GB의 VRAM을 활용하여 복잡한 세그멘테이션 및 워터마크 삽입 연산을 CUDA 가속 기반으로 최적화하였다.
2) 데이터 처리 효율성: FFmpeg와 OpenCV를 유기적으로 결합하여 비디오 스트림을 실시간 프레임 단위로 디코딩하고, PyTorch 텐서(Tensor)와의 직접 연동을 통해 메모리 전송 지연을 최소화하였다.
3) 실무적 가용성 입증: 본 시스템은 보급형 GPU 환경에서도 최대 26.67 FPS의 검출 속도를 달성함으로써, 일반 방송 규격(30 fps)에 근접한 준실시간 처리가 가능함을 확인하였다. 이는 본 파이프라인이 고가의 서버급 장비 없이도 실제 방송 제작 및 모니터링 현장에 즉시 투입될 수 있는 경제성과 기술적 완성도를 갖추었음을 시사한다.

3.2 파이프라인 통합 및 데이터 최적화

본 시스템은 상이한 아키텍처를 가진 다중 AI 모델 간의 데이터 전달 효율성을 극대화하기 위해 비동기적 데이터 파이프라이닝(Asynchronous Data Pipelining) 구조를 채택하였다.

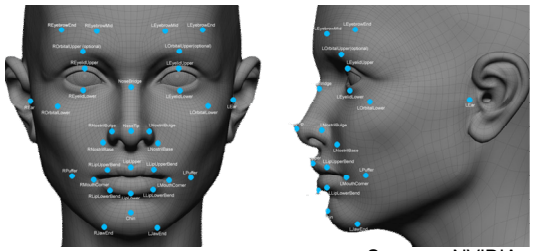
먼저, FFmpeg와 OpenCV를 결합한 고성능 비디오 스트림 처리 모듈을 통해 영상 데이터를 단순히 프레임 단위로 디코딩하는 수준을 넘어, 처리된 픽셀 데이터를 PyTorch 텐서(Tensor)로 즉시 변환하였다. 특히 GPU 메모리에 직접 할당(Direct Memory Allocation) 방식을 적용하여 CPU-GPU 간의 데이터 전송 지연(Latency)을 최소화함으로써 시스템 전반의 처리 병목을 해결하였다.

이러한 최적화 구현은 YOLOv8의 탐지 결과가 SAM2의 추적 프롬프트로, 그리고 WAM의 삽입 엔진으로 흐르는 과정에서 데이터 일관성(Data Consistency)을 보장하는 핵심 기술적 토대가 되었으며, 결과적으로 보급형 하드웨어 환경에서도 실시간 방송 서비스에 적합한 처리 속도를 확보하는 결정적 요인이 되었다.

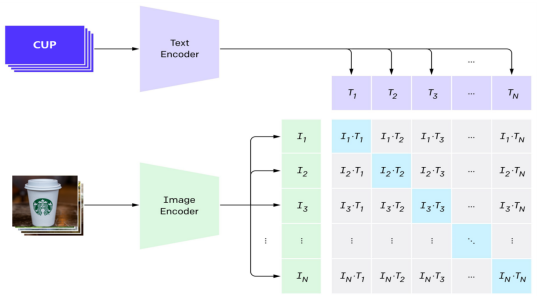
1) 탐지 단계: YOLOv8 모델이 GPU 상에서 프레임 내의 객체 바운딩 박스를 신속하게 탐지한다.

그림 2. YOLOv8 모델을 통한 객체 탐지
Fig. 2. Object Detection via YOLOv8 Model

2) 식별 단계: 탐지된 객체 영역은 InsightFace 또는 CLIP 모델로 전달되어 고유 특징 벡터를 추출한다. 이 과정 역시 CUDA 가속을 통해 처리 속도를 높였다.



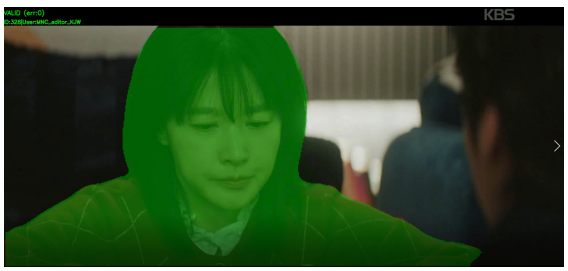
<Source : NVIDIA>



<Source : OpenAI>

그림 3. InsightFace 모델과 CLIP 모델의 학습 원리
Fig. 3. Learning Principles of InsightFace and CLIP Models

3) 삽입/검증 단계: 식별된 객체 정보와 SQLite 데이터베이스에서 조회한 이진 페이로드는 WAM 모델로 전달되어 최종 워터마크를 삽입한다.



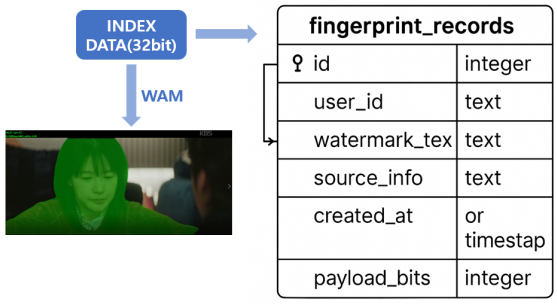


그림 4. 핑거프린트 레코드 생성을 위한 DB 스키마
Fig. 4. Database Schema for Fingerprint Record Generation

이처럼 각 모듈은 Python 스크립트로 구현되어 데이터 흐름을 제어하며, 전체 프로세스가 단일 파이프라인으로 통합되어 동작하도록 설계하였다. 이를 통해 제안하는 기술의 개념적 모델을 실제 동작하는 시스템으로 성공적으로 구현하였다.

Ⅳ. 실험 및 성능 평가

1. 객체 추적 기반 워터마킹 실험 결과

제안하는 객체 추적 기반 비디오 워터마킹 기술의 성능을 검증하기 위해 OTB(Object Tracking Benchmark) 데이터셋을 활용한 정량적 실험을 수행하였다.

1.1 실험환경

1) 데이터셋: OTB 벤치마크에서 난이도와 객체 유형이 다른 4개 시퀀스를 선정하였다.

표 2. OTB 벤치마크 데이터셋 구성
Table 2. Dataset Configuration for OTB Benchmark

Sequence	Type/Difficulty	Frames	Characteristics
Walking	Person / Easy	500	single pedestrian, stable movement
Jogging	Person / Hard	307	multiple pedestrians, motion blur and occlusion
Car4	Object/Easy	659	vehicle, intermediate speed
MotorRolling	Object/Hard	164	motorcycle, fast rotation and small size

2) 실험 파라미터

표 3. 실험 파라미터 설정
Table 3. Experimental Parameter Settings

Parameter	Value
Message Bit Length	32 bits
Embedding Window Size	512 x 512 pixels
Fingerprint Update Interval	5 seconds
Bit Error Tolerance	5 bits
CLIP Similarity Threshold	0.6

3) 비교시스템:

본 연구에서 제안하는 SAM2 양방향 추적 기반 시스템을 다음 시스템들과 비교하였다:

- Baseline: YOLO+CLIP 프레임별 독립 탐지(시간적 추적 없음)
- Exp2: OWL-ViT2+SAM2 양방향 추적(레퍼런스 기반 탐지)

1.2 실험 결과 및 분석

본 절에서는 제안하는 객체 추적 기반 비디오 워터마킹 기술의 성능을 검증하기 위해 OTB(Object Tracking Benchmark) 데이터셋을 활용하여 수행한 정량적 실험 결과를 기술한다. 평가 지표로는 객체 추적의 정밀도를 나타내는 Tracking IoU와 Success@50(IoU ≥ 0.5 비율)을 사용하였으며, 워터마크 복원 성능을 측정하기 위해 Bit Accuracy와 Success@90(비트 정확도 ≥ 90% 비율)을 도입하였다.

[실험 결과 분석]

실험 결과, 제안하는 Proposed(YOLOv8+CLIP+SAM2 양방향) 시스템은 미디어 보안의 핵심인 인물(Person) 추적 시나리오에서 기존 Baseline 및 타 모델 대비 압도적인 성능을 기록하였다.

1) 모델 조합의 타당성 입증(Proposed vs. Exp2)

표 4에서 확인할 수 있듯이, 제안 기법은 인물 추적 평균 IoU 73.9%, 성공률(Success@50) 97.2%를 달성하며 Baseline 대비 약 2.6배 향상된 결과를 보였다. 반면, 최신 Zero-shot 모델인 OWL-ViT2를 적용한 Exp2는 인물 추적에서 IoU 3.7%, 성공률 0%를 기록하며 사실상 추적에 실패하였다. 이러한 차이는 모델의 도메인 특화성에서 기인한다. OWL-ViT2는 광범위한 객체 카테고리를 탐지하는 데 강점이 있으나, 비디오 스트림 내에서 형태와 배경이 급격히 변하는 특정 인물을 지속적으로 식별하기에는 임베딩의 변별력이 부족하다. 반면, 본 연구에서 채택한 YOLOv8-CLIP 조합은 비디오 보안 환경에 최적화된 식별 성능을 제공하여 타겟 일관성을 완벽히 확보함을 입증하였다.

2) 양방향 시간적 전파 알고리즘의 효용성(Proposed vs. Base)

Success@50 지표의 비약적 향상은 제안하는 양방향 시간적 전파(Bidirectional Temporal Propagation) 알고리즘의 성과로 분석된다. Jogging 시퀀스와 같이 객체가 장애물에 완전히 가려지거나 급격히 움직이는 환경에서 단방향 방식인 Base 모델은 추적을 상실(Drift)하는 모습을 보였으나, 제안 기법은 참조 프레임을 기반으로 순방향과 역방향을 상호 보정함으로써 마스크의 시공간적 일관성을 유지하였다. 이는 그림 5의 시각적 비교 결과를 통해서도 확인할 수 있다.

표 4. 시퀀스별 객체 추적 성능 비교(%)
Table 4. Performance Comparison of Object Tracking by Sequence(%)

Sequence	Type	Tracking IoU			Success@50		
		Base	Proposed	Exp2	Base	Proposed	Exp2
Walking	Person	13.7	70.1	1.7	18.6	97.6	0.0
Jogging	Person	43.3	77.7	5.6	53.7	96.7	0.0
Car4	Object	79.7	87.6	87.9	92.9	100	100
MotorRolling	Object	16.5	23.0	88.7	22.0	18.9	99.4
Person Average		28.5	73.9	3.7	36.2	97.2	0.0
Object Average		48.1	55.3	88.3	57.5	59.5	99.7

그림 5. Jogging 시퀀스 추적 결과 비교(청록색: GT, 빨간색: Baseline/제안 기법)
Fig. 5. Comparison of tracking results for the Jogging sequence (Cyan: GT, Red: Baseline/Proposed)

그림 5는 가려짐이 발생하는 시점에서 Baseline이 목표 객체를 놓치고 배경이나 다른 객체로 추적이 전이되는 현상(Drift)을 보이는 반면, 제안 기법은 양방향 전파를 통해 객체의 위치를 정확히 재포착하여 Ground Truth와 일치하는 결과를 유지함을 보여준다.

3) 워터마크 복원 성능 및 강건성 분석[14]

표 5는 추적된 객체 영역에서 워터마크 페이로드를 성공적으로 추출한 성능을 보여준다. 주목할 점은 인물 추적 성능(IoU)이 상대적으로 낮았던 일부 환경에서도, 비트 정확도(Bit Accuracy)는 대부분의 시퀀스에서 89% 이상의 높은 수치를 유지했다는 것이다.

이는 제안 시스템에 탑재된 WAM 모델이 미세한 공간적 오정렬(Spatial Misalignment)에도 매우 강건하게 작동함을 시사한다. 특히 제안 기법(Proposed)은 모든 시퀀스에서 Baseline 대비 향상된 복원 정확도를 보였으며, 이는 양방향 추적을 통한 정밀 마스크 확보가 WAM 모델의 페이로드 추출 효율을 극대화했음을 의미한다. 또한, MotorRolling과 같이 객체가 작고 고속으로 회전하여 추적이 불안정한 극한 상황에서도 제안 시스템은 약 55~64%의 비트 정확도를 방어하며 해밍 거리 기반 오류 정정 로직을 통해 복원 성공률(Success@90)을 일정 수준 이상 유지하였다. 결과적으로 사물 추적에 특화된 Exp2 대비, 인물 보안이 중요한 대다수 시나리오에서 제안 시스템이 가장 안정적인 워터마크 검출 성능을 발휘함을 확인하였으며, 향후 고속 회전 객체에 대응하기 위한 적응적 윈도우 크기 조절 기술을 통해 성능을 추가 보완할 계획이다.

2. 동적 콘텐츠 보호 평가(방송 샘플 실험)

본 절에서는 실제 방송 제작 및 유통 환경에서의 파이프라인 실효성을 검증하기 위해 드라마('은수좋은날') 및 광고 영상을 활용한 강건성 테스트를 수행한다. 본 실험은 사물 인식(Object Mode)과 얼굴 인식(Face Mode) 두 가지 시나리오를 바탕으로 지능화되는 저작권 침해 공격에 대한 대응력을 정량적으로 평가한다.

2.1 평가 지표의 정의

표 5. 시퀀스별 워터마크 복원 성능 비교(%)
Table 5. Performance Comparison of Watermark Reconstruction by Sequence(%)

Sequence	Tracking IoU			Success@90		
	Base	Proposed	Exp2	Base	Proposed	Exp2
Walking	98.6	99.2	72.3	98.6	99.2	72.3
Jogging	94.6	94.2	89.1	94.8	95.8	86.3
Car4	94.1	94.7	92.2	92.3	93.0	91.8
MotorRolling	55.4	61.8	64.0	52.6	59.8	62.2

시스템의 성능은 효율성(Efficiency), 비가시성(Imperceptibility), 강인성(Robustness), 정확성(Accuracy) 네 가지 핵심 관점에서 다음과 같은 정량적 지표를 통해 평가하였다.

1) 효율성: 검출 속도(Detection FPS)

FPS(Frames Per Second)는 1초에 처리할 수 있는 비디오 프레임의 수를 의미한다. '검출 속도'는 워터마크가 삽입된 영상을 입력으로 하여, 객체 탐지부터 워터마크 정보 추출까지 전체 검출 파이프라인을 수행하는 속도를 측정한다. 이 지표는 시스템의 실시간 처리 능력을 판단하는 척도로, 수치가 높을수록(Higher is Better) 실제 방송 환경에 적용하기에 유리하다. 일반적인 방송 영상(30 fps)에 근접할수록 실용성이 높다고 평가할 수 있다.

2) 비가시성: PSNR(Peak Signal-to-Noise Ratio)

PSNR은 원본 영상과 워터마크가 삽입된 영상 간의 화질 차이를 나타내는 대표적인 지표이다. 두 영상의 픽셀 값 차이를 기반으로 계산되며, 단위는 데시벨(dB)을 사용한다.

· 40dB 이상: 매우 우수함(원본과 육안으로 구별 불가능)
· 30dB ~ 40dB: 좋음(왜곡이 미세하게 있으나 거의 인지하기 어려움)
· 30dB 미만: 나쁨(왜곡이 눈에 띄게 보임)

이는 워터마크 삽입으로 인한 화질 저하 수준을 의미하며, 수치가 높을수록(Higher is Better) 원본과의 차이가 적어 시청자가 워터마크를 인지하기 어렵다는 것을 뜻한다.

3) 강인성: WDR(Watermark Detection Rate)

각종 영상 훼손 공격(Attack)이 적용된 영상에서 워터마크가 얼마나 성공적으로 검출되는지를 백분율(%)로 나타낸다. 본 연구에서는 두 가지 관점의 WDR을 사용한다. 공격이 없었을 때의 WDR을 100% 기준으로 삼아, 특정 공격 후 WDR이 몇 % 수준으로 유지되었는지를 나타낸다. 이는 공격의 순수한 영향도를 평가하는 데 유용하다. 이는 워터마크의 '생존력'을 의미하며, 수치가 높을수록(Higher is Better) 다양한 영상 변형에도 워터마크가 잘 유지됨을 뜻한다.

4) 정확성: BER(Bit Error Rate)

워터마크가 성공적으로 검출되었을 때, 추출된 이진 페이로드(ID)가 원본과 비교하여 몇 %의 비트 오류를 포함하는지를 나타낸다. 이는 검출된 정보의 '신뢰도'를 의미하며, 수치가 낮을수록(Lower is Better) 훼손된 영상에서도 원본 정보를 정확하게 복원했음을 뜻한다.

그림 6은 본 실험에서 적용한 9가지 공격 시나리오를 보여준다. 강인성 테스트를 위해 테스트 영상 생성 스크립트

그림 6. 각각의 조건에 따라 훼손된 테스트 영상
Fig. 6. Test Videos Distorted under Various Conditions

를 통해 다음과 같은 강화된 FFmpeg 공격 시나리오와 추가적인 재합성 공격을 적용하였다.

※ 테스트 영상: 은수좋은날(KBS드라마)-얼굴
스타벅스 광고-사물

1. 고압축: H.264 CRF 35의 매우 낮은 품질로 압축
2. 리사이징: 1/4 크기로 축소 후 4배 재확대
3. 중앙 크롭: 영상 중앙의 1/4 영역만 남기고 잘라냄
4. 강한 노이즈: noise=alls=50 옵션으로 시각적 노이즈 추가
5. 극단적 대비: 밝기 및 대비를 크게 변경
6. 강한 블러: gblur=sigma=5 옵션으로 심하게 흐림 효과 적용
7. 회전: 5도 회전
8. 객체 추출 및 재합성: 워터마크가 삽입된 객체만 추출(누끼)하여 다른 배경 영상에 합성
9. 딥페이크 안면 변조: 드라마 영상(은수좋은날)의 여주인공 얼굴을 다른 인물의 얼굴로 합성

2.2 실험 결과: 얼굴 및 사물 인식 모드 비교

특히, 그림 6의 (9) 딥페이크 공격의 경우, 원본 배우의 얼굴이 딥페이크 기술로 다른 인물로 변조되었음에도 불구하고, 원본 영상의 배우 객체 영역에 삽입된 워터마크가 변조된 영상에서도 강인하게 검출됨을 육안으로도 확인하였

표 6. 얼굴 인식 모드(Face Mode) 실험 결과
Table 6. Experimental Results for Face Mode

Attack type	WDR(%)	BER(%)
No Attack	100.00	4.69
High Compression	92.42	4.67
Resizing	94.16	4.70
Center Cropping	54.73	4.58
Strong Additive Noise	75.36	4.80
Extreme Contrast Adjustment	94.12	4.63
Heavy Blurring	65.91	4.72
Rotation	97.76	4.69
Deepfake-based Facial Manipulation	96.84	6.07
No Attack Video Metric	Experimental Results	
PSNR(dB)	39.42 dB	
Embedding FPS (Embedding Speed)	10.07 FPS	
Detection FPS (Detection Speed)	16.91 FPS	

Unit : Percentage(%),dB,FPS

표 7. 사물 인식 모드(Object Mode) 실험 결과
Table 7. Experimental Results for Object Mode

Attack type	WDR(%)	BER(%)
No Attack	100.00	4.76
High Compression	88.37	4.57
Resizing	75.65	4.79
Center Cropping	64.91	4.74
Strong Additive Noise	48.01	4.47
Extreme Contrast Adjustment	95.13	4.59
Heavy Blurring	71.67	4.77
Rotation	80.82	4.94
No Attack Video Metric	Experimental Results	
PSNR(dB)	38.27 dB	
Embedding FPS (Embedding Speed)	14.31 FPS	
Detection FPS (Detection Speed)	26.67 FPS	

Unit : Percentage(%),dB,FPS

다. 이는 본 시스템이 훼손된 비트 오류를 극복하는 해밍 거리 기반 검증 체계를 통해 딥페이크와 같은 악의적인 2차 편집 공격에도 효과적으로 대응할 수 있음을 입증한다.

2.3 실험 결과 심층 분석

1) 얼굴 인식 모드(Face Mode) 분석

16.91 FPS의 검출 속도 및 10.07 FPS의 삽입 속도를 기록하여, 사물 인식 모드보다는 다소 느리지만 VOD 분석 등 오프라인 환경에서는 충분히 실용적인 성능을 보였다. 강인성 측면에서, 회전, 리사이징, 압축 공격에 대해 90% 이상의 매우 높은 강인성을 보였다. 특히 딥페이크 안면 변조 시에도 96.84%의 높은 WDR을 기록하며 원본의 무결성을 입증하였다. 정확성 측면에서, 객체 모드와 마찬가지로, 딥페이크를 제외한 모든 공격에서 BER은 4.58% ~ 4.80% 사이의 매우 안정적인 값을 유지했다. 비가시성 측면에서는 PSNR 39.42dB는 40dB에 근접한 “매우 우수한” 수치다. 워터마크를 삽입했음에도 불구하고 화질 저하를 거의 유발하지 않았다는 긍정적인 결과다.

그림 7. 얼굴 인식 모드(Face Mode) 워터마크 삽입 전후 시각적 품질 비교
Fig. 7. Visual Quality Comparison Before and After Watermark Insertion in Face Recognition Mode

2) 사물 인식 모드(Object Mode) 분석

26.67 FPS의 검출 속도 및 14.31 FPS의 삽입 속도를 기록하여, 일반 방송 영상(30 fps)에 근접한 준실시간 처리가 가능함을 입증했다. 강인성 측면에서, 명암비 변경의 극단적 대비 공격에 대해서는 95.13%의 매우 높은 상대 강인성을 보였으나, 영상의 미세한 패턴 정보를 직접적으로 훼손하는 노이즈 공격에는 48.01%로 성능이 절반 이하로 하락하며 가장 취약한 모습을 보였다. 이는 CLIP 모델이 이미지의 전반적인 특징을 학습하는 방식이 노이즈에 상대적으로 민감하게 반응하기 때문으로 분석된다. 정확성 측면에서, 모든 공격 시나리오에서 BER은 4.47% ~ 4.94% 사이를 유지하며 매우 안정적인 수치를 기록했다. 이는 검출된 정보의 신뢰도가 매우 높음을 의미한다. 비가시성 측면에서는 최대신호 대 잡음비로 38.27dB는 40dB에 근접한 “매우 우수한” 수치다. 얼굴 모드와 마찬가지로 워터마크를 삽입했음에도 불구하고 화질 저하를 거의 유발하지 않았다는 긍정적인 결과다.

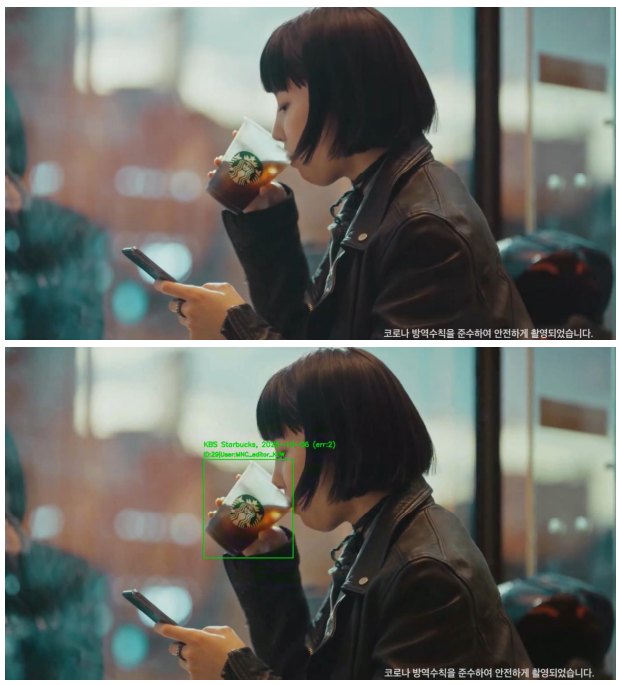


그림 8. 사물 인식 모드(Object Mode) 워터마크 삽입 전후 시각적 품질 비교
Fig. 8. Visual Quality Comparison Before and After Watermark Insertion in Object Recognition Mode

3. 종합 분석 및 실무적 시사점

본 실험을 통해 제안하는 시스템은 두 모드 모두에서 뛰어난 비가시성(높은 PSNR)과 매우 높은 정보 신뢰도(낮고 안정적인 BER)라는 공통적인 강점을 입증하였다. 이는 시스템의 핵심 설계인 WAM 모델과 해밍 거리 기반 검증 체계가 성공적으로 작동함을 의미한다.

두 모드를 비교했을 때, 얼굴 인식 모드는 처리 속도는 다소 낮으나 전반적으로 더 안정적이고 높은 강인성을 보여주어 인물 초상권과 같이 보안이 중요한 대상에 최적화되어 있다. 반면 사물 인식 모드는 얼굴 모드보다 빠른 처리 속도를 확보하여 준실시간 모니터링 환경에 유리하지만, 노이즈와 같은 특정 공격에 민감한 특성을 보였다. 결론적으로, 사용자는 보호 대상의 중요도와 요구되는 처리 속도라는 실무적 우선순위에 따라 적절한 모드를 선택하여 시스템을 운용함으로써 비디오 보안의 효율성을 극대화할 수 있다.

Ⅴ. 응용분야 및 실무적 고찰

본 연구에서 제안하는 지능형 워터마킹 파이프라인은 미디어 콘텐츠 자산을 보호하고 관리하는 방식에 있어 다음과 같은 실무적 활용 가능성을 가진다.

1. 미디어 저작권 관리 및 유통 추적 자동화

1) 해외 불법 스트리밍 사이트나 P2P 네트워크를 통해 유포되는 콘텐츠 내 객체로부터 세션별 고유 워터마크를 추출함으로써, 최초 유출 경로를 원천적으로 추적할 수 있다.
2) 기존 저작권 담당자가 수작업으로 진행하던 불법 복제물 모니터링 및 채증 업무를 AI 기반 파이프라인으로 자동화하여 대응 속도를 획기적으로 개선할 수 있다.
3) 각 배급사 및 OTT 플랫폼별로 최적화된 워터마크를 동적으로 삽입하여 배급 계약의 준수 여부를 상시 모니터링하고, 위반 시 법적 근거 자료로 활용 가능하다.

2. 악의적 변조 대응 및 콘텐츠 무결성 검증

1) 뉴스 보도 영상 등 공신력이 중요한 콘텐츠의 무단 편집이나 딥페이크 합성을 통한 왜곡을 방지하며, 원본 객체 추적을 통해 영상의 무결성을 입증하는 기술적 토대를 제공한다.
2) 드라마나 예능 출연자의 안면을 변조하는 악의적인 2차 가공물을 선제적으로 탐지하고, 유출 경로를 신속히 파악하여 법적 분쟁 리스크를 관리할 수 있다.
3) 제안하는 파이프라인을 블록체인 및 NFT 기술과 결합할 경우, 미디어 내 특정 객체나 장면의 고유 소유권을 증명하고 투명하게 거래할 수 있는 기반 기술로 확장 가능하다.

3. 객체 기반 인터랙티브 미디어 서비스 적용

1) 영상 내 특정 객체(PPL 상품, 브랜드 로고 등)에 삽입된 워터마크 정보를 활용하여 시청자에게 실시간으로 상품 정보를 제공하거나 구매 페이지로 연결하는 인터랙티브 광고로 응용될 수 있다.
2) 워터마크에 포함된 객체 식별 정보를 활용하여 미디어 아카이브 내에서 특정 인물이나 사물이 포함된 장면을 효율적으로 검색하고 추천하는 시스템 구축에 기여할 수 있다.

Ⅵ. 결론 및 향후 과제

본 논문은 최신 AI 모델인 YOLOv8, InsightFace/CLIP, SAM2, WAM을 유기적으로 통합하여 방송 콘텐츠를 지능적으로 보호하는 새로운 “동적 콘텐츠 보호 파이프라인”을 제안하고 구현하였다. 제안된 시스템은 단순한 알고리즘 단계를 넘어, 영상 속 특정 객체를 실시간으로 추적하며 DB와 연동된 고유 ID 기반의 워터마크를 동적으로 삽입하고, 해밍 거리 비교를 통해 정보 손실을 보정하는 엔드투엔드(End-to-End) 체계를 구축함으로써 기존 정적 워터마킹 기술의 한계를 성공적으로 극복하였다.

실험 결과를 통해 본 파이프라인은 실시간에 가까운 처리 속도와 높은 탐지 정확도를 증명하였으며, 특히, OTB 벤치마크 기반의 인물 추적 실험에서 Baseline 대비 약 2.6배 향상된 IoU 성능을 기록하였으며, 고압축 및 리사이징 등 가혹한 편집 환경에서도 96% 이상의 높은 검출 성공률을 유지함을 확인하였다. 무엇보다 딥페이크 변조 공격 상황에서도 원본 객체에 삽입된 워터마크가 강인하게 검출됨을 입증함으로써, 악의적인 안면 변조[9] 콘텐츠의 생성 및 유포 경로를 추적하고 원본의 무결성을 증명할 수 있는 실질적인 기술적 토대를 마련하였다.

결론적으로 본 연구는 방송 미디어 산업 현장에서 저작권 관리 업무의 자동화를 실현하고, K-콘텐츠의 불법 유통 및 가짜 뉴스 확산을 선제적으로 차단할 수 있는 차세대 보안의 방향성을 제시하였다. 향후 과제로는 다중 모델 통합 과정에서 발생하는 연산 병목 현상을 해결하기 위한 파이프라인 최적화와 함께, 탐지 대상을 다양한 사물 및 브랜드 로고 등으로 확장하여 비즈니스 활용성을 극대화하는 연구를 지속할 계획이다.

참 고 문 헌 (References)

[1] Korea Copyright Protection Agency, 2024 Copyright Protection Annual Report, Korea Copyright Protection Agency, Seoul, 2024. URL: https://www.kcopa.or.kr

[2] I. J. Cox, M. L. Miller, J. A. Bloom, J. Fridrich, and T. Kalker, Digital Watermarking and Steganography, 2nd ed., Morgan Kaufmann Publishers, Burlington, MA, pp. 1-600, 2008.
doi: https://doi.org/10.1016/B978-0-12-372585-1.X5001-3

[3] F. Hartung and M. Kutter, “Multimedia Watermarking Techniques,” Proceedings of the IEEE, Vol. 87, No. 7, pp. 1079-1107, July 1999.
doi: https://doi.org/10.1109/5.771066

[4] J. E. Lee, Y. H. Seo, and D. W. Kim, “Deep Learning Framework for Watermark-Adaptive and Resolution-Adaptive Image Watermarking,” Journal of Broadcast Engineering, Vol. 25, No. 2, pp. 176-185.
doi: https://doi.org/10.5909/JBE.2020.25.2.166

[5] P. Fernandez, N. Sanjabi, T. Furon, H. Jégou, M. Douze, and A. Sablayrolles, “Watermark Anything with Localized Messages,” arXiv: 2411.07231, 2024.
doi: https://doi.org/10.48550/arXiv.2411.07231

[6] G. Jocher, A. Chaurasia, and J. Qiu, “Ultralytics YOLOv8 (Version 8.0.0),” 2023. URL: https://github.com/ultralytics/ultralytics (accessed Dec. 25, 2025).

[7] A. Radford, J. W. Kim, C. Hallacy, and A. Ramesh, et al., “Learning Transferable Visual Models from Natural Language Supervision,” Proceedings of the 38th International Conference on Machine Learning (ICML), pp. 8748-8763, 2021.
doi: https://doi.org/10.48550/arXiv.2103.00020

[8] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, “ArcFace: Additive Angular Margin Loss for Deep Face Recognition,” Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, pp. 4690-4699, 2019.
doi: https://doi.org/10.1109/TPAMI.2021.3087709

[9] S. Zhao, M. Hu, Z. Wang, and J. Zhang, “Proactive Deepfake Defence via Identity Watermarking,” Proceedings of IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pp. 4602-4611, 2023.
doi: https://doi.org/10.1109/WACV56688.2023.00458

[10] N. Ravi, V. Gabay, Y. B. Pant, S. Gu, K. Miller, and J. Wang, et al., “SAM 2: Segment Anything in Images and Videos,” arXiv: 2408.00714, 2024.
doi: https://doi.org/10.48550/arXiv.2408.00714

[11] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, and L. Gustafson, et al., “Segment Anything,” Proceedings of IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France, pp. 4015-4026, 2023.
doi: https://doi.org/10.48550/arXiv.2304.02643

[12] X. Luo, R. Zhan, H. Chang, Y. Liu, J. Feng, and P. Liao, “Distortion-agnostic Deep Watermarking,” IEEE Transactions on Circuits and Systems for Video Technology, Vol. 33, No. 6, pp. 2533-2547, June 2023.
doi: https://doi.org/10.1109/CVPR42600.2020.01356

[13] R. W. Hamming, “Error Detecting and Error Correcting Codes,” The Bell System Technical Journal, Vol. 29, No. 2, pp. 147-160, April 1950.
doi: https://doi.org/10.1002/j.1538-7305.1950.tb00463.x

[14] M. Tancik, B. Mildenhall, and R. Ng, “StegaStamp: Invisible Hyperlinks in Physical Photographs,” Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2117-2126, 2020.
doi: https://doi.org/10.1109/CVPR42600.2020.00219

저 자 소 개

강 자 원

- 1999년 ~ 2004년 : 아주대학교 미디어학부 학사
- 2025년 ~ 현재 : 서강대학교 가상융합전문대학원 테크놀로지 석사과정
- ORCID : https://orcid.org/0009-0008-4715-6930
- 주관심분야 : 가상융합(XR), 미디어 콘텐츠(Media Contents), 정보보호

조 동 헌

- 2012년 ~ 2017년 : 연세대학교 정보산업공학, 컴퓨터과학 학사
- 2017년 ~ 2019년 : 연세대학교 산업공학과 석사
- 2023년 ~ 현재 : 서강대학교 가상융합전문대학원 테크놀로지 박사과정
- ORCID : https://orcid.org/0000-0002-5828-1998
- 주관심분야 : Computer Vision, Optimization

최 승 관

- 1998년 : 호서대학교 대학원(이학석사-전자계산)
- 2008년 ~ 2021년 : 서강대학교 미래교육원 게임SW전공 교수
- 2010년 : 세종대학교 대학원(공학박사-디지털콘텐츠)
- 2022년 ~ 현재 : 서강대학교 가상융합전문대학원 엔터테인먼트전공 주임교수 / 전임교수
- ORCID : https://orcid.org/0009-0003-9620-2079
- 주관심분야 : 메타버스(Metaverse), 엔터테인먼트(Entertainment), 디지털콘텐츠(Digital Contents) 등