The Transactions of the Korea Information Processing Society
TKIPS Vol.13, No.10, pp.481~491

eISSN: 3022-7011
https://doi.org/10.3745/TKIPS.2024.13.10.481

Hierarchical Watermarking Technique Combining Error Correction Codes

Do-Eun Kim† · So-Hyun Park†† · Il-Gu Lee†††

ABSTRACT

Digital watermarking is a technique for embedding information into digital content. Digital watermarking has attracted attention as a technique to combat piracy and identify artificially generated content, but it is still not robust in various situations. In this paper, we propose a frequency conversion-based hierarchical watermarking technique capable of attack detection, error correction, and owner identification. By embedding attack detection and error correction signatures in hierarchical watermarking, the proposed scheme maintains invisibility and outperforms the existing methods in capacity and robustness. We also proposed a framework to evaluate the performance of the image quality and error correction according to the type of error correction signature and the number of signature embeddings. We compared the visual quality and error correction performance of the conventional model without error correction signature and the conventional model with hamming and BCH signatures. We compared the quality by the number of signature embeddings and found that the quality deteriorates as the number of embeddings increases but is robust to attacks. By analyzing the quality and error correction ability by error correction signature type, we found that hamming codes showed better error correction performance than BCH codes and 41.31% better signature restoration performance than conventional methods.

Keywords : Digital Watermarking, Copyright, Security, Error Correction Code, Noise Attack, Otsu Algorithm

오류 정정 부호를 결합한 계층적 워터마킹 기법

김 도 은† · 박 소 현†† · 이 일 구†††

요 약

디지털 워터마킹은 디지털 컨텐츠에 정보를 삽입하는 기술이다. 불법 복제 근절과 인공지능이 생성한 콘텐츠 식별 기술로 디지털 워터마킹이 주목받고 있지만, 여전히 다양한 상황에서 견고하지 못하다. 본 논문에서는 공격 탐지와 오류 정정 및 소유자 식별이 가능한 주파수 변환 기반의 계층적 워터마킹 기법을 제안한다. 제안 방식은 계층적 워터마킹에 공격 탐지 및 오류 정정 시그니처 삽입을 통해 비가시성을 유지하며 용량과 견고성 측면에서 종래 방법보다 향상된 성능을 보였다. 또한 오류정정부호 종류와 시그니처 삽입 횟수에 따른 이미지 품질과 오류정정 성능을 비교 평가하는 프레임워크를 제안하여, 오류정정부호가 없는 종래 모델과 해밍 부호 및 BCH 부호를 적용한 종래 모델의 시각적 품질과 오류정정 성능을 비교 평가하였다. 시그니처 삽입 횟수에 따른 품질을 비교한 결과에 따르면 삽입 횟수가 증가할수록 품질은 열화되나 공격 상황에 견고하였다. 그리고 오류정정부호 종류에 따른 품질과 오류 정정 능력을 분석한 결과에 따르면 BCH 부호보다 해밍 부호 사용 시 향상된 오류정정 성능을 보였으며, 종래 방식 대비 41.31% 향상된 시그니처 복원 성능을 보였다.

키워드 : 디지털 워터마킹, 저작권, 보안, 오류정정부호, 노이즈 공격, Otsu 알고리즘

1. 서 론

디지털 경제(Digital Economy) 시대에 콘텐츠의 생산, 유통, 소비 활동의 장벽은 유례없이 낮아졌다. 이미지 처리 및 생산 기술과 인터넷의 대중화로 디지털 컨텐츠를 저렴한 비용으로 누구나 쉽게 수정, 복제, 재생산 및 배포할 수 있게 되었다. 스마트 기기와 소셜 네트워킹 서비스의 일상화로 디지털 콘텐츠의 경제·사회적 파급력은 꾸준히 높아질 것으로 예상된다. 그러나 콘텐츠의 유통과 소비가 간편해짐에 따라 무단 복제 및 배포 등 콘텐츠의 불법 도용이 성행하게 되었으며, 저

※ 이 논문은 2024년도 산업통상자원부 및 한국산업기술진흥원의 산업혁신인재성장지원사업 (RS-2024-00415520)과 과학기술정보통신부 및 정보통신기획평가원의 ICT혁신인재4.0 사업의 연구결과로 수행되었음(No. IITP-2022-RS-2022-00156310).
※ 이 논문은 2024년 ASK 2024의 우수논문으로 “디지털 워터마킹 공격 탐지를 위한 계층적 워터마킹 기법”의 제목으로 발표된 논문을 확장한 것임.
† 비 회 원 : 성신여자대학교 융합보안공학과 학사과정
†† 준 회 원 : 성신여자대학교 미래융합기술공학과 박사과정
††† 종신회원 : 성신여자대학교 융합보안공학과/미래융합기술공학과 부교수
Manuscript Received : July 11, 2024
Accepted : September 6, 2024
* Corresponding Author : Il-Gu Lee(iglee@sungshin.ac.kr)

작권 침해와 데이터 위변조 문제가 심각한 사회·경제적 이슈로 떠올랐다[1, 2]. 그리고 최근에는 불법 복제를 근절하는 것뿐만 아니라 인공지능이 생성한 콘텐츠 식별이 디지털 콘텐츠 산업의 중요한 과제가 되었다[3]. 최근 생성형 인공지능 기술에 대한 접근성의 확장으로, 인공지능이 생성하는 디지털 컨텐츠의 공급량이 증가하고 있다. 한국은 딥페이크를 활용한 가짜 뉴스 등의 위험에 대응하기 위해 인공지능 생성물에 대한 표식을 의무화하도록 했으며, 미국과 EU 역시 국가 차원에서 관련 제도를 마련하고 있다[4, 5]. 디지털 정보 산업에 중요성이 커지고 있는 지적 자산(Intellectual Property)인 디지털 콘텐츠 창작물의 권리 보호와 창작 주체의 식별이 뒷받침되지 않는다면, 창작 활동에 대한 적절한 수익 분배가 불가능하여 디지털 콘텐츠 시장이 왜곡될 수 있다. 소셜 네트워크 서비스의 일상화와 생성형 인공지능의 대중화로 디지털 콘텐츠의 생산량이 빠르게 증가하는 현시점에서, 디지털 워터마킹(Digital Watermarking)은 저작권 및 소유권 증명 및 보호 추적 기술로써 콘텐츠 산업의 성장과 여러 파생 시장의 존립을 결정하는 중요한 기술이다[6]. 디지털 워터마킹은 신뢰할 만한 창작 환경을 조성하고 창작물 수익에 대한 정당한 수익 분배에 기여하여 창작 활동을 보호하고 활성화하는 핵심적인 인프라로 주목받고 있다.

디지털 워터마킹(Digital Watermarking)은 이미지·음향·영상 등의 디지털 컨텐츠에 저작권 정보를 삽입하여 저작권 및 소유권을 보호함으로써 불법 복제를 방지한다[7]. 종래의 디지털 워터마킹의 알고리즘은 호스트 이미지의 공간을 분할하여 정보를 LSB(Least Significant Bit) 픽셀에 직접 삽입하는 공간 도메인(Spatial Domain) 방식과 DWT(Discrete wavelet transform), DCT(Discrete Cosine Transform), SVD(Singular Value Decomposition)와 같이 호스트 이미지를 주파수 도메인으로 변환하여 부가 정보를 삽입하는 주파수 도메인 방식이 대표적이다[8]. 또한 디지털 워터마킹 기술의 주요 속성으로는 비가시성(Invisibility), 용량(Capacity), 견고성(Robustness)이 있다. 비가시성은 디지털 워터마킹 시스템의 성능을 측정하는 주요 지표로서 워터마크가 적용된 이미지는 사람이 인지할 수 없는 품질을 유지해야 한다. 용량은 원본 호스트 이미지에 삽입할 수 있는 정보의 양을 평가하는 것이며, 견고성이란 워터마킹 시스템이 일반적 신호 처리 상황에서 간섭이나 변형을 받아도 워터마크를 감지하는지 평가하기 위한 요구사항이다[7]. 그러나 세 속성은 트레이드오프 관계를 가지므로 다른 요소를 저해하지 않으며 성능을 높이는 것은 연구자들이 당면한 과제이다. 또한, 디지털 워터마킹을 가우시안 노이즈 공격, 스펙클 노이즈 공격과 같은 노이즈 공격, 압축, 자르기, 변형, 적대적 공격 등의 기하학적 또는 비기하학적 공격에 취약하다[9].

이러한 문제점을 해결하기 위해 견고성과 저작권 보호를 고려한 연구가 활발히 이어지고 있다[18, 19]. 기존의 단일 알고리즘을 이용한 디지털 워터마킹 알고리즘에서 나아가 이산 웨이블릿 변환과 특이값 분해를 결합한 하이브리드 워터마킹 기법을 제안하여 비가시성을 보장하면서 공격에 견고한 하이브리드 워터마킹 기법을 제안하였다[10]. 최근에는 헬스케어 등의 분야에서 이미지 워터마킹을 통해 의료 이미지에 메타데이터를 삽입하거나 NFT(Non-Fungible Token) 기반 데이터 마켓 플레이스에서 디지털 워터마킹을 이용해 소유자와 구매자 정보를 삽입하는 등 워터마킹 기술을 다방면에 응용하는 연구도 진행되고 있다[11, 12].

본 논문에서는 오류정정부호를 활용한 계층적 워터마킹 기법을 제안한다. 디지털 컨텐츠의 저작권 및 소유권을 침해하는 공격에 대응하기 위해 오류정정부호를 활용하여 정보 주체의 디지털 ID를 인코딩하고 이를 이미지에 반복 삽입한 시그니처 이미지와 워터마크 이미지를 계층적으로 삽입 및 복원 추출하는 방식을 제안한다.

본 논문의 주요 기여점은 다음과 같다.

- 공격 및 훼손 상황에서 정보를 복원함으로써 저작권과 소유권을 보장 및 추적하기 위해 오류정정부호를 결합한 계층적 워터마킹 기법을 제안한다.
- 오류정정부호 종류와 시그니처 삽입 횟수에 따른 이미지 품질과 오류정정 성능을 비교 평가하는 프레임워크를 제안한다.
- 제안한 방식은 비가시성을 유지하며 용량과 견고성 측면에서 종래 방법보다 향상된 성능을 보였다.

본 논문의 2장에서는 계층적 삽입 및 추출 알고리즘을 제안하고, 3장에서 제안한 방식의 비가시성과 다양한 공격 상황에서의 견고성 및 오류정정 성능을 평가한다. 그리고 4장에서 결론을 맺는다.

2. 오류정정부호를 결합한 계층적 워터마킹

본 논문에서 제안한 계층적 워터마킹은 호스트 이미지를 주파수 대역별로 분할한 뒤, 두 개의 부대역에 두 이미지를 삽입 및 추출하는 알고리즘이다. 주파수를 분할하여 워터마크 이미지와 워터마크 위변조 공격 탐지를 위한 시그니처를 각각 삽입함으로써 용량 효율적으로 이미지 워터마킹 및 공격 탐지 기능을 수행할 수 있다. 또한, 오류정정부호를 결합한 계층적 워터마킹은 디지털 콘텐츠의 소유자를 식별하고 임베딩한 워터마크 훼손 공격을 탐지하기 위하여 정보 주체의 Digital ID를 오류정정부호를 이용하여 인코딩하고, 이미지 형태의 공격 탐지 시그니처를 생성한다. 이 방식은 비가시성을 보장하면서도 오류정정 기능을 통해 공격 상황에서 견고성을 보인다. 워터마크 이미지와 시그니처 이미지의 삽입 알고리즘으로는 DWT(Discrete wavelet transform)를 이용하였다.

2.1 계층적 삽입 알고리즘

Fig. 1과 Algorithm 1은 제안하는 계층적 삽입 방식의 동작 흐름도와 알고리즘을 보여준다.

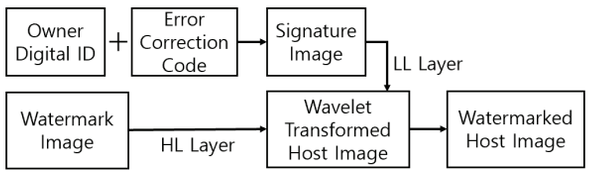

Fig. 1. Flowchart of Hierarchical Insertion Algorithm

Algorithm 1
Pseudocode of the Hierarchical Insertion Algorithm

Input: DID (Owner's Digital ID), Host Image H, Watermark Image W
Output: Watermarked Host Image H'
DID_binary = convert_to_binary(DID)
Signature = encode_ECC(DID_binary)
Procedure create_signature_image(ECC)
    Initialize Signature_Image as empty
    For i = 1 **to** r **do**
        Append ECC to Signature_Image
        Append padding to Signature_Image
    return Signature_Image
Procedure Hierarchical_Insertion(H, S, W)
    Discrete Wavelet Transform to H
    Embed Signature Image into LL band
    Embed Watermark Image into HL band
    return the Watermarked Host Image H'

먼저, 정보 주체 및 소유자의 디지털 ID를 생성하고 2진수로 변환한다. 2진수로 변환된 디지털 ID를 이용해 오류정정부호를 생성하고, 오류정정부호의 2진수 시퀀스를 이미지의 흑/백 픽셀로 치환하여 padding과 함께 반복 인코딩하는 방식으로 공격 탐지 시그니처 이미지를 생성한다. 본 연구에서는 8자리 10진수 디지털 ID를 이용하였으며, 각 10진수를 4비트의 2진수로 변환하여, 32bit의 디지털 ID로 오류정정부호를 생성하고 이미지에 반복 인코딩하였다. 사용한 오류정정부호는 해밍 부호(Hamming codes)와 BCH 부호(BCH codes, Bose-Chaudhuri-Hocquenghem codes)이다. 이후, Haar-wavelet을 적용해 이산 웨이블릿 변환한 호스트 이미지에 위변조 탐지를 위한 시그니처 이미지를 저주파(LL, low-low) 대역에 삽입하고, 워터마크 로고 이미지를 중주파(HL, high-low) 대역에 삽입한다.

한 번의 이산 웨이블릿 변환을 통해 분할한 주파수 대역에 계층별로 두 가지 이미지를 임베딩 함으로써 호스트 이미지 내의 제한된 공간에 더 많은 정보를 임베딩할 수 있으므로 용량 측면에서 자원을 효율적으로 활용할 수 있다.

2.2 계층적 추출 알고리즘

계층적 워터마킹의 추출 알고리즘은 Fig. 2와 Algorithm 2와 같다. 워터마킹된 호스트 이미지를 이산 웨이블릿 변환한 뒤 HL 대역과 LL 대역에서 삽입한 이미지를 추출한다. HL 대역에서는 추가적인 처리 과정 없이도 추출한 워터마크 이미지와 원본 워터마크 이미지와의 NCC(Normalized Cross-Correlation)값이 우수한 품질의 이미지를 추출할 수 있다. LL 대역에서 추출한 시그니처 이미지에서, 시그니처를 복원하고 추출하기 위해 오류정정을 수행하고 흑/백 픽셀값을 이진수 0/1으로 변환하여 공격 탐지 시그니처를 추출한다. 제안하는 오류정정 기능을 갖춘 계층적 추출 방식은 공격 및 워터마크 훼손 상황에서 공격 탐지 시그니처를 오류정정 디코딩하여 훼손된 시그니처 정보와 사용자의 디지털 ID를 복원할 수 있다.

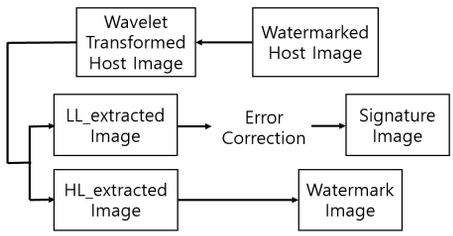

Fig. 2. Hierarchical Extraction Algorithm

Algorithm 2
Pseudocode of the Hierarchical Extraction Algorithm

Input: Watermarked Host Image H'
Output: Signature_restored, Extracted Signature Image S_extracted, Extracted Watermark Image W_extracted
Procedure Hierarchical_Extraction(H')
    Discrete Wavelet Transform to H'
    S_extracted = Extract Image from LL band
    W_extracted = Extract Image from HL band
    return S_extracted, W_extracted
Procedure Error_Correction(S_extracted)
    ECC_corrected = error_correction(S_extracted)
    Signature_binary = convert_to_binary(ECC_corrected)
    Signature_restored = binary_to_signature(Signature_binary)
    return Signature_restored
Decoding Signature_restored to DID

3. 성능 평가 및 분석

본 연구에서는 제안한 방식과 종래 방법의 성능을 평가하기 위해 512x512 사이즈의 이미지들을 활용하여 실험했다. 호스트 이미지로 Lenna를 이용했고, 워터마크 이미지로는 Pepper 이미지를 이용하였다. Intel(R) Core(TM) i7-1360P CPU @ 2.20 GHz, Intel(R) Iris(R) Xe Graphics, RAM 16G, Python 3.11.5 버전 환경에서 실험하였다.

시그니처 이미지는 오류정정부호로 인코딩한 사용자의 디지털 ID를 Table 1과 같이 300, 600, 900회 반복 삽입하여 생성하였다. 또한 오류정정부호의 성능 평가를 위해 오류정정부호를 이용하지 않은 uncoded와 1bit의 오류를 정정할 수 있는 해밍 부호 및 5bit의 오류를 정정하도록 구현한 BCH 부호를 이용한 경우를 비교하였다.

Table 1. Encoding Types and Size of Signature Images

Encoding Type	Available Error Correction [bit]	Signature Repetition [회]	Digital ID size [bit]	ECC Signature size [bit]	Padding size [bit]
uncoded	-	300	32	32	840
		600	32	32	400
		900	32	32	256
hamming	1	300	32	38	830
		600	32	38	396
		900	32	38	252
BCH	5	300	32	72	800
		600	32	72	396
		900	32	72	252

제안하는 계층적 워터마킹 삽입 알고리즘을 이용해 시그니처 이미지와 워터마크 이미지를 삽입한 Watermarked Host Image를 생성하고, 이를 대상으로 비공격, 가우시안 노이즈 공격(Gaussian noise attack), 스펙클 노이즈 공격(Speckle noise attack)을 수행한 뒤, HL 대역에서 추출한 워터마크 이미지와 LL 대역에서 추출한 시그니처 이미지의 품질과 오류 정정 성능을 평가하였다. 호스트 이미지의 경우 시각적 유사도인 PSNR(Peak Signal-to-Noise Ratio)과 구조적 유사도인 SSIM(Structural Similarity Index Map)으로 비가시성을 평가하였고, HL 대역에서 추출한 워터마크의 품질은 NCC, LL 대역에서 추출한 시그니처 이미지 품질은 NCC와 BER(Bit Error Rate)로 측정하였다.

먼저, 종래 단일 이미지 삽입 워터마킹 방식과 제안하는 계층적 워터마킹 방식의 삽입 강도에 따른 비가시성을 비교 분석하였다. 주파수 분할 기반 워터마킹 구현을 위해 Haar-wavelet을 적용한 DWT로 호스트 이미지를 주파수 대역별로 분할한 상태에서 단일 및 계층적 워터마크 삽입이 이루어졌다. Fig. 3은 종래의 단일 삽입 워터마킹 방식과 제안하는 계층적 워터마킹 방식 구현에 사용한 이미지이다.

단일 삽입 워터마킹은 HL 대역에 워터마크 이미지 b)를 삽입하였으며, 제안하는 계층적 워터마킹 방식은 HL 대역에 워터마크 이미지 b)와 LL 대역에 시그니처 이미지 c)를 삽입하였다. c)는 해밍 부호로 디지털 ID를 인코딩한 값을 300회 반복 삽입한 시그니처 이미지이며 d)는 호스트 이미지를 주파수 대역 분할하여 워터마크 이미지와 시그니처 이미지를 계층적으로 삽입한 제안한 모델의 결과 이미지이다.

단일 삽입 워터마킹은 HL 대역의 삽입 강도를 0.01부터 0.1까지 조정하였고, 계층적 삽입 워터마킹은 HL 대역의 삽입 강도를 0.05로 고정하고 LL대역의 삽입 강도를 0.01부터 0.1까지 조정하였다. 삽입 강도에 따른 단일 삽입 및 계층적 워터마킹을 수행한 호스트 이미지의 시각적 유사도인 PSNR과 구조적 유사도 SSIM을 측정한 결과는 Fig. 4와 같다.

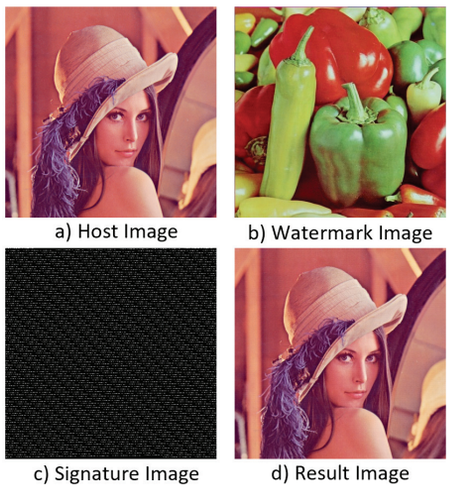

Fig. 3. a) Host Image, b) Watermark Image, c) Signature Image, d) Result Image

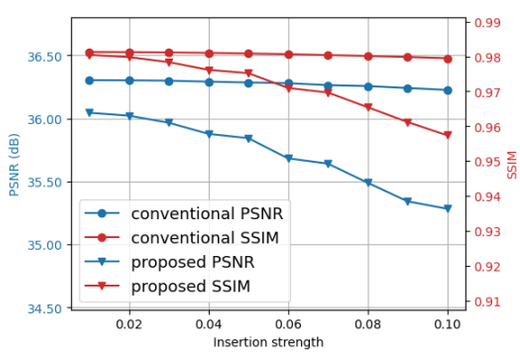

Fig. 4. SSIM and PSNR per Watermarking Type and Insertion Strength

종래의 단일 이미지 워터마킹의 경우 삽입 강도를 0.1으로 높인 경우에도 삽입 결과 호스트 이미지의 PSNR이 36.23dB, SSIM은 0.9795로 높은 비가시성을 보인다. 제안하는 계층적 워터마킹 방식에 따라 시그니처 이미지를 삽입했을 때 종래의 단일 이미지 워터마킹 방식 대비 비가시성이 감소한다. 그러나 HL 대역 삽입 강도 0.05 상황에서 LL 대역 삽입 강도를 0.1까지 높인 경우에도 결과 호스트 이미지의 PSNR이 35.28dB, SSIM이 0.9573로 시각적, 구조적 비가시성이 우수하다.

시그니처 이미지의 시그니처 반복 횟수와 오류정정부호 유무 및 종류를 변화하며 계층적 워터마킹 구현 후 비공격, Gaussian noise 공격, Speckle noise 공격 상황에서 Watermarked Host Image와 HL 대역과 LL 대역에서 추출한 워터마크 이미지와 시그니처 이미지의 품질과 오류 정정 성능을 분석한 결과는 다음과 같다.

3.1 비공격 상황

본 연구에서는 HL 대역에 단일 이미지를 삽입한 호스트 이

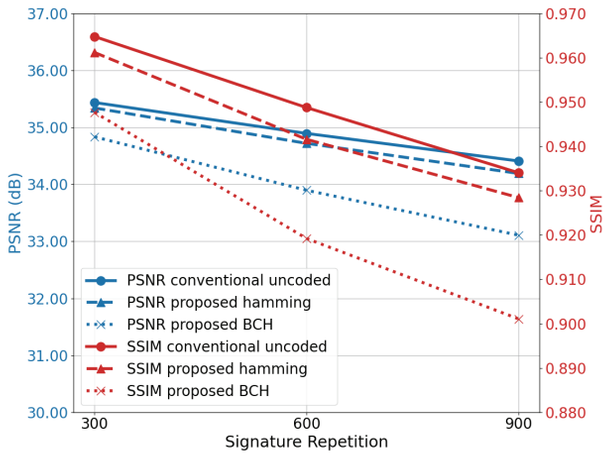

Fig. 5. SSIM and PSNR per Signature Repetition

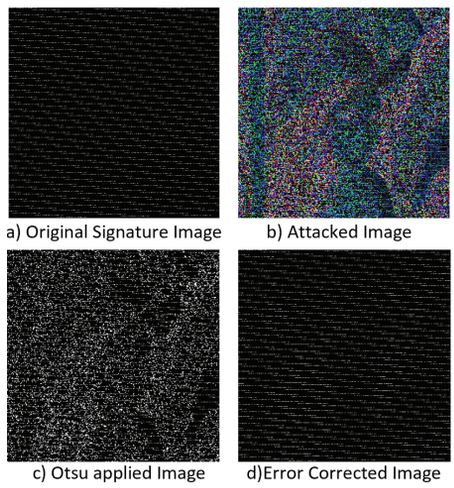

Fig. 6. a) Original Signature Image, b) Attacked Image, c) Otsu applied Image, d) Error Corrected Image

미지의 비가시성을 측정한 Fig. 3의 환경 조건에서 HL 대역 삽입 강도를 0.05, LL 대역 삽입 강도를 0.09로 설정하였다. Haar-wavelet을 적용해 주파수 대역별로 이산 웨이블릿 변환한 호스트 이미지의 HL 대역에는 워터마크 이미지를 삽입했고, LL 대역에는 시그니처 이미지를 삽입하였다. 삽입한 시그니처 이미지의 시그니처 반복 횟수와 시그니처의 오류정정부호 유무 및 종류에 따른 결과 이미지 PSNR과 SSIM은 Fig. 5와 같다.

Fig. 5에서 비교한 모든 경우는 높은 비가시성을 보장하는 PSNR과 SSIM 값을 보이고 있지만, 시그니처 반복 삽입 횟수와 오류정정부호 유무 및 종류에 따라 차이를 보인다. 공격 탐지 시그니처를 300, 600, 900회 반복 삽입할 때 종래 방식과 제안하는 오류정정부호 기반 계층적 워터마킹 방식의 PSNR과 SSIM을 비교하였고, 반복 삽입 횟수가 늘어날수록 워터마킹 삽입 결과 이미지의 PSNR과 SSIM이 낮아진다.

또한, 오류정정부호의 유무에 따라 오류정정 기능 없이 디지털 ID만을 반복 삽입한 기존 방식인 uncoded와 오류정정부호를 결합한 제안 방식을 비교하였다. 실험 환경에서 해밍부호는 1bit, BCH 부호는 5bit의 오류 정정 능력을 가지며, 오류정정 인코딩을 수행하지 않은 uncoded는 32bit, 해밍 부호로 인코딩한 시그니처의 길이는 38bit, 그리고 BCH 부호로 인코딩된 시그니처는 72bit의 길이를 가진다. 실험 결과에 따르면 워터마크에 삽입하는 시그니처의 길이가 길어질수록 결과 이미지의 PSNR과 SSIM이 낮아지는 것을 확인할 수 있다.

HL 대역에 워터마크 이미지를 삽입하고 LL 대역에 시그니처 이미지를 삽입하는 계층적 삽입 결과에 따르면 시그니처 이미지에 삽입한 정보 비트 수가 많을수록 워터마킹을 삽입한 이미지의 구조적 유사도인 SSIM과 시각적 유사도인 PSNR이 낮아진다. LL 대역에 삽입한 시그니처 이미지에서 공격 탐지 시그니처의 반복 삽입 횟수가 증가할수록, 오류정정부호 인코딩 길이가 길수록 인코딩하는 비트 수가 증가하여 계층적 삽입한 이미지의 비가시성은 열화되었다.

3.2 가우시안 노이즈 공격

가우시안 노이즈 공격이란 세마포어 노이즈의 일종으로, 정규분포 함수를 따르는 무작위 값을 신호나 이미지에 추가하는 가우시안 노이즈 모델을 이용한 공격이다[13]. 본 연구에서는 노이즈의 정규분포 파라미터를 조절하여 가우시안 노이즈 공격 강도에 따른 훼손 정도와 시그니처 이미지의 오류정정 능력을 분석하였다.

Fig. 6은 표준편차 0.5인 가우시안 노이즈 공격을 수행한 상황의 시그니처 이미지 상태 변화이다. a)는 정보 주체의 디지털 ID를 BCH 부호 기반으로 인코딩한 값을 300회 반복 삽입한 시그니처 이미지이고, b)는 가우시안 노이즈의 표준편차를 0.5로 설정하고 공격을 수행한 후 추출한 시그니처 이미지이다. c)는 b)에서 Otsu 알고리즘을 적용한 이미지이며, d)는 b)에서 BCH 오류 정정을 수행한 이미지이다.

Otsu 알고리즘이란 이미지의 임계값을 계산하여 사전에 결정된 클래스로 분류하는 이미지 처리 기법이다. Otsu 알고리즘은 그레이 레벨 이미지 강도를 세분화하여 L개의 레벨로 분류하고, 레벨별로 존재하는 픽셀 수를 히스토그램으로 나타낸 뒤, 이를 확률 분포로 간주하고 임계값을 설정한다. Otsu 알고리즘은 일반적으로 흑과 백으로 구성된 2개의 클래스를 이용하여 그레이 레벨 이미지를 두 픽셀값으로 이진화하는 데 이용한다[14]. Fig. 6에서 가우시안 노이즈 공격 직후 추출한 b) 이미지는 픽셀의 R(Red), G(Green), B(Blue)값으로 인해 이진수 시그니처를 추출하지 못한다. 이때 Otsu 알고리즘을 적용하면 Fig. 6의 c)와 같은 이진화된 이미지를 얻어 이진수 시그니처를 추출할 수 있다.

1) 중주파 대역 분석

가우시안 정규분포의 표준편차와 시그니처 이미지 조절에 따른 가우시안 노이즈 공격 후 HL 대역에서 추출한 워터마크 이미지의 NCC값은 Fig. 7과 같다. 가우시안 노이즈의 표준편

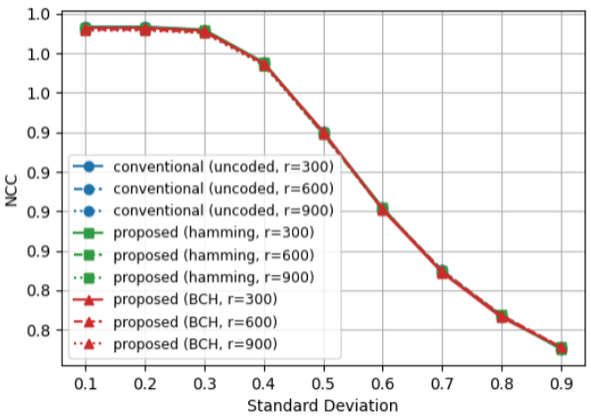

Fig. 7. HL Layer Extracted Watermark Image NCC per Standard Deviation

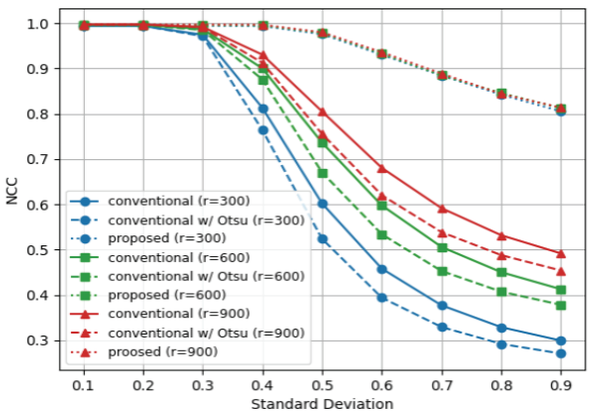

Fig. 8. LL Layer Extracted Signature Image NCC per Signature Repetition and Error Correction

차를 0.1부터 0.9까지 늘리며 공격을 수행하였고, 계층적 워터마킹에 삽입한 시그니처 이미지가 공격받은 HL 대역 워터마크 이미지에 미치는 영향을 평가하였다. 종래 방식과 제안하는 방식 모두 시그니처를 300, 600, 900회 반복 삽입한 횟수 r에 따른 NCC 값을 비교하였다. 또한 오류정정부호의 유무에 따라 오류정정 기능 없이 디지털 ID만을 반복 삽입한 기존 방식인 uncoded와 오류정정부호를 결합한 제안 방식을 비교하였으며, 오류정정부호는 1bit 오류정정을 수행하는 해밍부호, 5bit 오류정정을 수행하는 BCH 부호를 이용하였다.

가우시안 정규분포의 표준편차 증가에 따라 노이즈가 증가하여 HL 대역에서 추출한 워터마크 이미지의 NCC 값은 점차 감소하며 품질이 떨어졌다. 그러나 HL 대역의 워터마크 이미지 추출 품질은 LL 대역에 삽입한 시그니처 이미지의 시그니처 반복 삽입 횟수와 오류정정부호의 종류에 영향을 받지 않아 유사한 수준의 품질 저하를 보인다.

2) 저주파 대역 분석

Fig. 8, 9, 10은 시그니처 이미지의 시그니처 반복 횟수와 오류정정부호의 유무 및 종류를 조정하며 가우시안 노이즈 공격을 수행했다. 이후 공격 강도에 따라 LL 대역에서 추출한 시그니처 이미지의 품질과 오류정정 능력을 비교 분석하였다.

Fig. 8은 시그니처 삽입 횟수와 오류정정 기능 유무에 따른 가우시안 노이즈 공격 시 시그니처 이미지의 품질을 나타낸 것이다. BCH 오류정정부호를 인코딩한 시그니처를 삽입 횟수 r에 따라 300, 600, 900회 삽입한 시그니처 이미지와 워터마크 이미지를 계층적 워터마킹 삽입한 호스트 이미지에 가우시안 노이즈 공격을 수행하였다. 이후 이미지 처리 방식별로 LL 대역에서 추출한 시그니처 이미지의 NCC를 측정하였다.

conventional은 가우시안 노이즈 공격 직후 추출한 시그니처 이미지, conventional w/ Otsu는 conventional에서 시그니처 추출을 위해 종래의 Otsu 알고리즘을 적용한 이미지, proposed은 conventional에서 오류정정을 수행한 제안모델이다. Fig. 8 그래프에서 conventional은 Fig. 6의 b), conventional w/ Otsu는 Fig. 6의 c), proposed은 Fig. 6의 d)와 같은 상황이다. 공격받은 이미지를 Otsu 알고리즘으로 이진화해 추출한 시그니처와 오류정정부호를 이용해 오류를 복구해 추출한 시그니처의 오류율을 비교하였다.

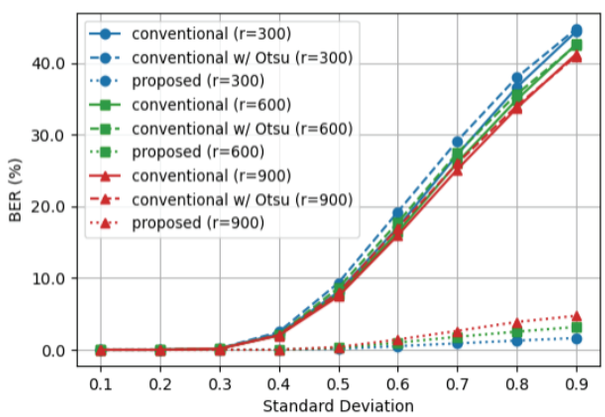

Fig. 9. LL Layer Extracted Signature Image BER per Signature Repetition and Standard Deviation

실험 결과에 따르면 가우시안 노이즈 공격의 표준편차 증가에 따라 시그니처 이미지의 NCC 값이 저하된다. 훼손된 시그니처는 BCH 오류정정으로 복원하여 우수한 NCC를 얻을 수 있으며, 시그니처 삽입 횟수와 관계없이 유사한 수준으로 NCC가 증가한다. 예를 들어, 표준편차 0.9, 시그니처 300회 삽입한 conventional (r=300)의 경우 공격 직후 LL 대역에서 추출한 시그니처 이미지와 원본 시그니처 이미지의 NCC 값이 0.2993으로 심각하게 훼손되었다. Otsu 알고리즘을 이용하면 공격받은 이미지를 이진화하여 2진수 시그니처를 추출할 수 있지만, conventional w/ Otsu (r=300)의 NCC는 0.2707로 감소하여 conventional (r=300) 보다 품질이 낮아진다. 반면, 제안 방식으로 훼손된 BCH 코드를 추출하여 오류정정을 수행한 proposed (r=300)은 NCC를 0.8051까지 높일 수 있으며, 표준편차를 0.9까지 높인 상황에서도 300회 인코딩한 시그니처 중 284회 원본 시그니처를 추출할 수 있었다.

Fig. 9는 Fig. 8과 같은 조건에서 추출한 시그니처 이미지의 BER을 측정한 결과이다. BER은 비트 오류율으로, 통신 시스템에서 송수신 데이터 차이를 측정하는 지표로 이용된다. 본 절에서는 원본 시그니처 이미지와 공격 후 추출한 LL 대역의 시그니처 이미지 사이의 픽셀 값 차이를 측정하는 지표로 이용하였다.

Equation (1)은 비트 오류율을 측정한 방식이다.

$$BER(\%) = \frac{N_e}{N_i} \times 100 \quad (1)$$

N_i는 이미지의 전체 비트 수이며, N_e는 원본 이미지 중에서 에러가 발생한 비트 수이다.

가우시안 노이즈 공격의 표준편차 증가에 따라 시그니처 이미지의 BER이 증가하고, 공격 직후인 conventional보다 conventional w/ Otsu 알고리즘을 적용했을 때 BER이 근소하게 높아졌다. 그러나 BCH 오류정정을 수행하면 BER을 대폭 낮출 수 있으며, 이때 시그니처 삽입 횟수가 BER 복원 성능에 미치는 영향은 미미하다. 예를 들어, 표준 편차를 0.9, 시그니처를 300회 인코딩한 conventional (r=300)의 경우 공격 직후 LL 대역에서 추출한 시그니처 이미지와 원본 시그니처 이미지와의 BER은 44.41%으로 매우 훼손된 상태이다. conventional (r=300) 이미지에 Otsu 알고리즘을 통해 이진화한 conventional w/ Otsu (r=300)의 BER은 44.77%으로 오류율이 근소하게 증가한다. 본 논문의 제안 방식을 따르는 BCH 오류 정정을 수행한 proposed (r=300)의 BER은 01.63%로 우수한 오류 정정 성능을 보인다.

Fig. 8과 Fig. 9에서 제안하는 계층적 추출 알고리즘에 따라 BCH 오류 정정을 수행하면 우수한 NCC와 BER 값을 보이는 시그니처 이미지로 복원할 수 있다. 시그니처 반복 삽입 횟수가 늘어날수록 공격 후 추출한 NCC는 증가하고 BER은 약간 감소하는 양상을 보인다. NCC는 두 이미지의 픽셀값 간의 상관관계를 통해 유사성을 측정하기 때문에 반복 삽입한 시그니처의 패턴이 일부 유지되면 높은 NCC가 나올 수 있다[15]. 같은 노이즈 상황에서 900회 삽입한 시그니처 이미지가 300회 삽입한 시그니처 이미지보다 반복 삽입 패턴이 비교적 잘 유지되어 NCC가 높게 산출된다. 반면에, BER은 두 이미지의 픽셀값 차이를 직접적으로 비교하기 때문에 이미지의 작은 노이즈나 변화에도 민감하다. 900회 삽입한 시그니처 이미지는 300회 삽입한 시그니처 이미지보다 노이즈 환경에서 오염되지 않은 시그니처의 비율이 높아서 BER이 근소하게 감소한다.

Otsu 알고리즘 적용 여부인 conventional와 conventional w/ Otsu에서 시그니처 삽입 횟수에 따라 추출 이미지의 NCC, BER 값을 비교 분석한 결과에 따르면 시그니처 중복 삽입 횟수가 늘어날수록 공격 상황에서 추출한 LL 대역 이미지의 NCC 값이 증가한다. 또한 가우시안 노이즈가 Otsu 알고리즘의 최적 임계값 설정을 방해하여 Otsu 알고리즘을 적용한 conventional w/ Otsu는 공격 직후인 w/o Otsu 이미지보다 낮은 NCC 값을 보인다[16].

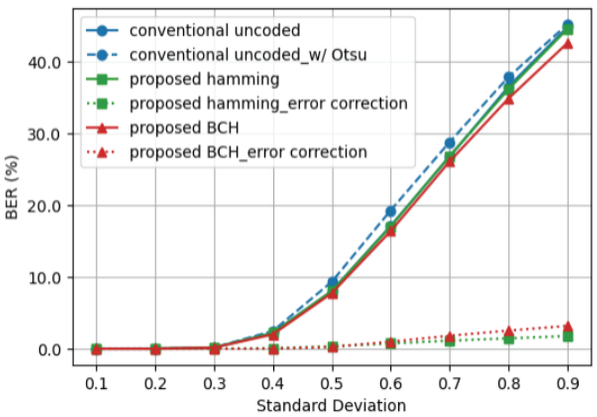

Fig. 10. LL Layer Extracted Signature Image BER per Error Correction Code and Standard Deviation

Fig. 10은 오류정정부호 유무 및 종류별로 uncoded, 해밍부호, BCH 부호를 적용한 시그니처를 600회 반복 삽입한 시그니처 이미지를 이용해 계층적 워터마킹한 이미지에 가우시안 노이즈 공격을 수행하였다. conventional uncoded, proposed hamming, proposed BCH는 공격 직후 추출한 시그니처 이미지이고, w/ Otsu는 훼손된 시그니처 이미지에 Otsu 알고리즘을 적용한 종래모델, 오류정정을 수행한 error correction은 제안 모델이다. 삽입 시 이용한 오류정정부호에 따른 시그니처 이미지의 BER을 측정한 결과, 인코딩한 부호의 유무 및 종류에 관계없이 노이즈 확률 증가에 따라 노이즈 주입량이 증가하여 BER은 점차 높아진다. 표준편차 0.9 상황에서 종래의 uncoded는 오류 정정 기능이 없어서 BER이 44.86%에 그쳤으나, 해밍 오류 정정 수행 시 BER이 44.60%에서 1.77%까지 감소하였고, BCH 오류 정정 수행 시 BER이 42.64%에서 3.18%까지 감소하였다.

해밍 부호와 BCH 부호의 경우, 표준편차 0.6 이상에서 해밍 부호가 더 우세하였다. 표준편차 0.9 상황에서 해밍 부호의 BER은 1.77%, BCH 부호의 BER은 3.18%로 해밍 부호가 1.7배 우수한 BER을 보였다. 노이즈 양의 증가로 오류를 정정할 수 없는 시퀀스가 증가하기 때문에 BCH 부호보다 인코딩 결과 비트 길이가 짧은 해밍 부호가 노이즈의 영향을 적게 받아서 BER이 낮게 나타나는 것이다. 본 실험에서 시그니처를 600회 반복 삽입 시, 72bit인 BCH 부호는 38bit인 해밍 부호에 비해 1.89배 많은 20,400bit를 추가로 삽입하게 되므로, 다중 오류정정 방식임에도 정보 픽셀이 가우시안 노이즈에 더 많이 오염되어 정정할 수 있는 비트 수를 초과하기 때문에 해밍 부호를 이용한 경우보다 높은 BER을 가진다.

3.3 스펙클 공격

스펙클 노이즈(Speckle Noise)는 활성 레이더 및 SAR (Synthetic Aperture Radar) 이미지의 품질을 저하하는 세분화된 노이즈이다[17]. 본 실험에서는 이미지의 각 픽셀에 스펙

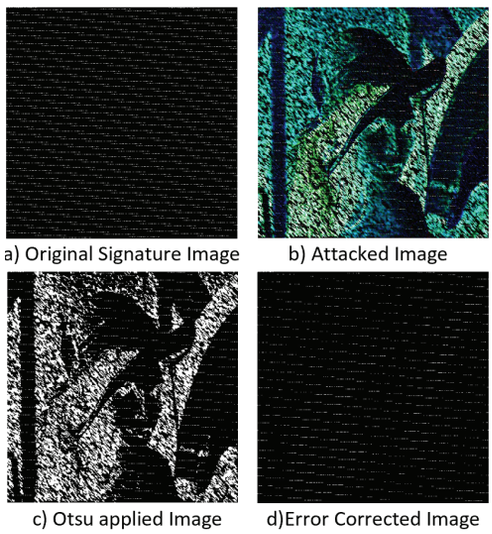

Fig. 11. a) Original Signature Image, b) Attacked Image, c) Otsu applied Image, d) Error Corrected Image

클 노이즈가 주입될 확률(Probability)에 따른 스펙클 공격을 수행하였다.

Fig. 11은 확률 0.05인 스펙클 노이즈 공격 상황의 시그니처 이미지 상태 변화이다. a)는 정보 주체의 디지털 ID로 생성한 해밍 부호를 300회 반복 인코딩한 시그니처 이미지이고, b)는 스펙클 노이즈 표준편차 0.5의 공격 수행 후 추출한 시그니처 이미지이다. c)는 b)에서 Otsu 알고리즘을 적용한 이미지이며, d)는 b)에서 해밍 오류 정정을 수행한 이미지이다.

1) 중주파 대역 분석

Fig. 12는 노이즈 주입 확률과 시그니처 이미지 조절에 따른 스펙클 노이즈 공격 후 HL 대역에서 추출한 워터마크 이미지의 NCC를 나타낸다. 스펙클 노이즈의 확률을 0.01부터 0.09까지 늘리며 공격을 수행하였고, 계층적 워터마킹에 삽입한 시그니처 이미지가 공격받은 HL 대역 워터마크 이미지에 미치는 영향을 평가하였다. 시그니처 이미지는 시그니처를 300, 600, 900회 반복 삽입한 횟수 r에 따라 차이를 두었다.

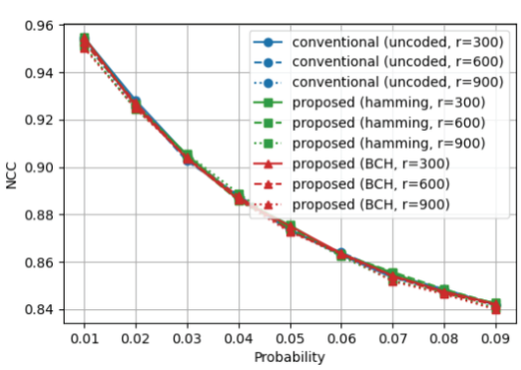

Fig. 12. HL Layer Extracted Watermark Image NCC per Probability

또한 오류정정부호의 유무에 따라 오류정정 기능 없이 디지털 ID만을 반복 삽입한 기존 방식인 uncoded와 오류정정부호를 결합한 제안 방식을 비교하였으며, 오류정정부호는 그 종류에 따라 1bit 오류정정을 수행하는 해밍 부호, 5bit 오류정정을 수행하는 BCH 부호를 이용하였다.

스펙클 노이즈 주입 확률 증가에 따라 HL 대역에서 추출한 워터마크 이미지의 NCC는 감소하며 품질이 떨어졌다. 그러나 HL 대역의 워터마크 이미지 추출 품질은 LL 대역에 삽입한 시그니처 이미지의 시그니처 반복 삽입 횟수와 오류정정부호의 종류에 영향을 받지 않아서 유사한 수준의 품질 저하를 보인다.

2) 저주파 대역 분석

Fig. 13은 시그니처 삽입 횟수와 오류정정 기능에 따른 스펙클 노이즈 공격 시 시그니처 이미지의 품질을 나타낸 것이다. BCH 부호를 인코딩한 시그니처를 삽입 횟수 r에 따라 300, 600, 900회 삽입한 시그니처 이미지와 워터마크 이미지를 계층적 워터마킹 삽입한 호스트 이미지에 스펙클 노이즈 공격을 수행하였다. 공격 후 이미지 처리 방식별로 LL 대역에서 추출한 시그니처 이미지의 NCC를 측정 및 분석하였다.

conventional은 스펙클 노이즈 공격 직후 추출한 시그니처 이미지, conventional w/ Otsu는 conventional에서 시그니처 추출을 위해 종래의 Otsu 알고리즘을 적용한 이미지, proposed는 conventional에서 오류정정을 수행한 제안모델이다. Fig. 13 그래프에서 conventional은 Fig. 11의 b), conventional w/ Otsu는 Fig. 11의 c), proposed는 Fig. 11의 d) 상황이다. 공격받은 이미지를 종래의 Otsu 알고리즘으로 이진화해 추출한 시그니처와 제안하는 오류정정부호를 이용해 오류를 복구하여 추출한 시그니처의 오류율을 비교했다.

실험 결과에 따르면 노이즈 주입 확률 증가에 따라 시그니처 이미지의 NCC 값이 저하된다. 훼손된 시그니처는 BCH 오류정정으로 복원하여 우수한 NCC를 얻을 수 있으며, 시그니처 삽입 횟수와 관계없이 유사한 수준으로 NCC가 증가한다.

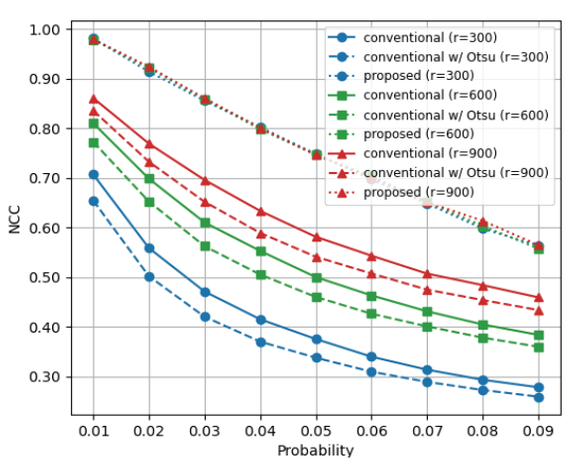

Fig. 13. LL Layer Extracted Signature Image NCC per Signature Repetition and Probability

예를 들어, 확률 0.05로 공격 직후 LL 대역에서 추출한 시그니처 이미지 conventional (r=300)는 원본 시그니처 이미지와의 NCC가 0.5000으로 심각하게 훼손되었다. Otsu 알고리즘을 이용하면 시그니처를 추출할 수 있지만, conventional w/ Otsu (r=300)의 NCC는 0.4599로 근소하게 낮아져 공격 직후보다 낮은 품질을 보인다. 반면 공격 직후 이미지인 conventional에 BCH 오류정정을 수행하면 원본 시그니처 이미지와의 유사도인 NCC를 0.7466까지 높일 수 있다.

Fig. 14는 Fig. 13과 같은 조건에서 추출한 시그니처 이미지의 BER을 보여준다. 스펙클 노이즈 공격의 확률 증가에 따라 시그니처 이미지의 BER이 증가하고, 공격 직후인 conventional보다 Otsu 알고리즘을 적용한 경우가 BER이 근소하게 높다. 그러나 BCH 오류정정을 수행하면 BER을 대폭 낮출 수 있다. 예를 들어, 확률 0.05, 시그니처 삽입 횟수 300인 conventional (r=300)은 공격 직후 LL 대역에서 추출한 시그니처 이미지의 NCC가 18.95%으로 심하게 훼손된 상태이다. conventional (r=300)에 Otsu 알고리즘을 적용해 이진화한 BER은 22.09%로 오류율이 조금 증가한다. 본 논문의 제안 방식을 따르는 proposed (r=300)를 적용하면 BER을 1.60%까지 낮추는 우수한 오류 정정 성능을 보인다.

Fig. 14와 Fig. 13에서 제안하는 계층적 추출 알고리즘에 따라 BCH 오류정정을 수행하면 우수한 NCC와 BER 값을 보이는 시그니처 이미지로 복원할 수 있다. 그러나 가우시안 노이즈 공격 상황과 동일하게 시그니처 반복 삽입 횟수가 증가할수록 NCC가 증가하고 BER은 감소한다. conventional 이미지와 conventional w/ Otsu 이미지 또한 가우시안 노이즈 공격 상황과 동일하게 시그니처 중복 삽입 횟수가 늘어날수록 NCC 값이 증가하였고. Ostu 알고리즘의 임계값 설정을 방해하는 노이즈 공격으로 인해 conventional w/ Otsu 이미지는 공격 직후의 conventional 이미지보다 낮은 NCC, 높은 BER 값을 보인다.

Fig. 15는 오류정정부호 유무 및 종류별로 uncoded, 해밍 부호, BCH 부호를 적용한 시그니처를 600회 반복 삽입한 시그니처 이미지를 이용해 계층적 워터마킹한 이미지에 스펙클 노이즈 공격을 수행하였다. conventional uncoded, proposed hamming, proposed BCH는 공격 직후 추출한 시그니처 이미지이고, w/ Otsu는 훼손된 시그니처 이미지에 Otsu 알고리즘을 적용한 종래모델, 오류정정을 수행한 error correction은 제안모델이다. 삽입 시 이용한 오류정정부호에 따른 시그니처 이미지의 BER을 측정한 결과에 따르면 인코딩한 부호의 유무 및 종류에 관계없이 노이즈 확률 증가에 따라 노이즈 주입량이 증가하여 BER이 커진다. 확률 0.09, 시그니처를 600회 반복 삽입하면 오류정정 기능이 없는 종래 uncoded 방식의 BER은 31.44%이다. 그러나 해밍 오류 정정 시 BER이 31.81%에서 2.77%까지 감소하였고, BCH 오류 정정 시 BER이 30.77%에서 5.11%까지 감소하여 원본 시그니처 이미지와 오류율을 낮출 수 있다. BCH 부호보다 해밍 부호가 더 낮은 BER로 복원하였다.

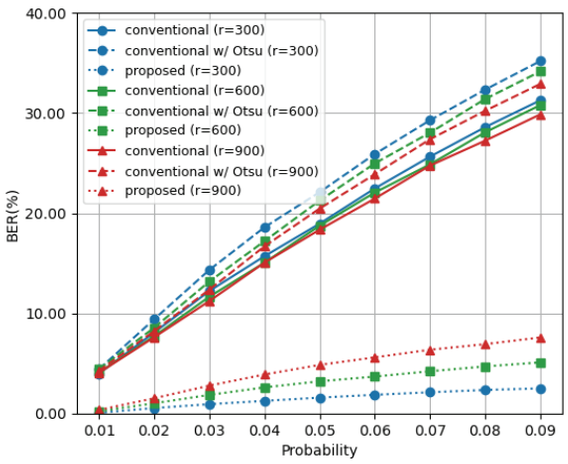

Fig. 14. LL Layer Extracted Signature Image BER per Signature Repetition and Probability

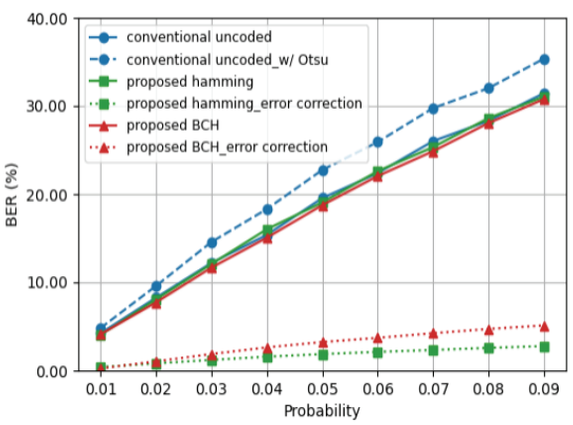

Fig. 15. LL Layer Extracted Signature Image BER per Error Correction Code and Probability

3.4 결과 분석

본 연구의 결과를 정리하면 다음과 같다. 오류정정부호 종류와 시그니처의 삽입 횟수를 변경하면 호스트 이미지의 비가시성에 영향을 미친다. 시그니처 삽입 횟수가 증가할수록, 인코딩 길이가 긴 오류정정부호 방식일수록 삽입하는 정보 길이가 늘어나서 생성한 이미지의 비가시성이 열화되지만, 제안 방식은 비가시성을 유지한다.

가우시안 노이즈 공격과 스펙클 노이즈 공격을 수행했을 때 HL 대역에서 추출한 워터마크 이미지의 NCC는 공격 강도의 증가에 따라 감소하고, 이때 LL 대역에 삽입한 시그니처 이미지의 오류정정부호 종류나 시그니처 반복 횟수는 결과에 영향을 미치지 않는다.

두 공격을 수행한 후 LL 대역에서 추출한 시그니처 이미지 또한 공격 강도가 증가할수록 품질이 훼손됨을 NCC와 BER로 확인하였다. 시그니처 이미지는 시그니처 삽입 횟수가 증가할수록 시그니처 패턴이 견고했고, 오류정정부호 길이가 짧을수록 시그니처가 받는 노이즈 영향이 줄어들어서 NCC는 증가하고 BER은 감소했다.

공격받은 LL 대역의 시그니처 이미지로부터 시그니처를 추출하기 위해, 복원 기능 없이 Otsu 알고리즘으로 이진화하여 시그니처를 추출하는 종래 방식과 오류정정 기능을 수행하여 복원 추출하는 제안 방식을 비교하였다. Otsu 알고리즘은 노이즈가 최적의 임계값 설정을 방해하여 성능이 저하되었지만, 오류정정부호를 통해 오류정정을 수행하면 높은 NCC와 낮은 BER의 시그니처 이미지를 복원할 수 있었다. 또한 시그니처 삽입 횟수가 복원 능력에 미치는 영향은 저조하며, 오류정정부호 길이로 인해 해밍 부호의 오류정정 성능이 BCH 부호보다 우수함을 BER로 확인하였다.

결론적으로, 호스트 이미지의 주파수 대역을 분할하고 HL 대역에 워터마크 이미지와 LL 대역에 시그니처 이미지를 삽입하면, 워터마크 훼손 공격 상황에서 시그니처 이미지를 통해 호스트 이미지에 대한 공격을 탐지할 수 있다. 또한, 시그니처 이미지를 오류정정 인코딩하면 노이즈 공격으로 인해 시그니처 이미지가 훼손되는 상황에서도 시그니처 이미지 복원이 가능하여 종래 방식 대비 더 높은 강인성을 보장할 수 있다. 실험 결과에 따르면 해밍 부호를 900회 삽입했을 때 가장 오류정정 성능이 뛰어났다.

4. 결 론

본 연구는 디지털 워터마크의 소유자 식별, 위변조 감지 및 오류정정 기능을 갖춘 계층적 워터마킹 방식을 제안하였다. 제안하는 삽입 알고리즘은 사용자의 오류정정부호로 디지털 ID를 인코딩한 시그니처를 반복 삽입해 소유자 식별 및 위변조 감지 기능을 하는 시그니처 이미지를 생성한 후, 워터마크 이미지와 시그니처 이미지를 호스트 이미지에 주파수 대역별로 삽입한다. 제안하는 추출 알고리즘은 워터마크 이미지가 훼손되는 공격 및 강한 노이즈 상황에서 오류정정부호로 오류를 정정하여 시그니처를 복원할 수 있다. 가우시안 노이즈 공격과 스펙클 노이즈 공격 상황에서 오류정정부호와 시그니처 삽입 횟수에 따른 이미지 품질과 오류정정 성능을 비교 평가하였다. 오류 복구 성능은 시그니처 삽입 횟수가 증가할수록 향상되며, 종래 방식의 44.23% BER을 2.64%까지 낮춘 해밍 부호가 가장 우세한 것으로 나타났다. 제안 방식은 워터마크 무력화 공격에 대한 견고성을 보장하는 동시에 비가시성을 유지하며 주파수 분할 기반 계층적 워터마킹으로 용량 효율성을 높일 수 있다. 본 연구에서는 블록 코딩 방식만을 다루었다. 향후 연구에서는 블록 코딩 및 컨볼루션 코딩 방식과 적용 방법 따른 오류정정 성능과 견고성을 분석하고자 한다.

References

[1] A. Ray and S. Roy, “Recent trends in image watermarking techniques for copyright protection: a survey,” *International Journal of Multimedia Information Retrieval*, Vol.9, No.4, pp.249-270, 2020.

[2] H. R. Hasan, K. Salah, R. Jayaraman, I. Yaqoob, and M. Omar, “NFTs for combating deepfakes and fake metaverse digital contents,” *Internet of Things*, Vol.25, p.101133, 2024.

[3] M. Fenwick and P. Jurcys, “Originality and the Future of Copyright in an Age of Generative AI,” *Computer Law & Security Review*, Vol.51, pp.105892, 2023.

[4] J. Kim, D. Im, and H. Cha, “Legislative Study on Directions Concerning Regulation of Generative Artificial Intelligence (AI) Systems - Focusing on Duty to Disclosure and Transparency Regarding the Creations of Generative AI Systems Such as ChatGPT,” *Contemporary Review of Criminal Law*, Vol.80, pp.245-283, 2023.

[5] C. Novelli, F. Casolari, P. Hacker, G. Spedicato, and L. Floridi, “Generative AI in EU law: liability, privacy, intellectual property, and cybersecurity,” *arXiv preprint arXiv: 2401.07348*, 2024.

[6] S. Wadhera, D. Kamra, A. Rajpal, A. Jain, and V. Jain, “A comprehensive review on digital image watermarking,” *arXiv preprint arXiv:2207.06909*, 2022.

[7] O. Evsutin and K. Dzhanashia, “Watermarking schemes for digital images: Robustness overview,” *Signal Processing: Image Communication*, Vol.100, pp.116523, 2022.

[8] A. Mohanarathinam, S. Kamalraj, G. K. D. Prasanna Venkatesan, R. V. Ravi, and C. S. Manikandababu, “Digital watermarking techniques for image security: a review,” *Journal of Ambient Intelligence and Humanized Computing*, Vol.11, No.8, pp.3221-3229, 2020.

[9] M. Begum and M. S. Uddin, “Digital Image Watermarking Techniques: A Review,” *Information*, Vol.11, No.2, pp.110, 2020.

[10] S. M. Arora, “A DWT-SVD based robust digital watermarking for digital images,” *Procedia Computer Science*, Vol.132, pp.1441-1448, 2018.

[11] G. Coatrieux, L. Lecornu, B. Sankur, and C. Roux, “A Review of Image Watermarking Applications in Healthcare,” in *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, pp.4691-4694, 2006.

[12] S. Ranjbar Alvar, M. Akbari, D. Yue, and Y. Zhang, “NFT-Based Data Marketplace with Digital Watermarking,” in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp.4756-4767, 2023.

[13] J. Yu, "Based on Gaussian filter to improve the effect of the images in Gaussian noise and pepper noise," *Journal of Physics: Conference Series*, Vol.2580, No.1, pp.012062, 2023.
[14] T. Y. Goh, S. N. Basah, H. Yazid, M. J. A. Safar, and F. S. A. Saad, "Performance analysis of image thresholding: Otsu technique," *Measurement*, Vol.114, pp.298-307, 2018.
[15] N. S. Hashemi, R. B. Aghdam, A. S. B. Ghiasi, and P. Fatemi, "Template matching advances and applications in image analysis," *arXiv preprint arXiv:1610.07231*, 2016.
[16] Q. Cao, L. Qingge, and P. Yang, "Performance Analysis of Otsu-Based Thresholding Algorithms: A Comparative Study," *Journal of Sensors*, Vol.2021. No.1, pp.4896853, 2021.
[17] A. Joseph and K. Anusudha, "Robust watermarking based on DWT SVD," *International Journal on Signal & Image Security*, Vol.1, No.1, pp.1-5, 2013.
[18] D. Kim, S. Park, and I. Lee, "Hierarchical watermarking technique for detecting digital watermarking attacks," *Proceedings of the Annual Symposium of Korea Information Processing Society Conference (KIPS)*, Vol.31, No.1, pp.283-284, 2024.
[19] S. Kim, Y. Jeon, and I. Lee, "Robust digital watermarking technique against distortion attacks," *Proceedings of the Annual Symposium of Korea Information Processing Society Conference (KIPS)*, Vol.31, No.1, pp.345-346, 2024.

김 도 은

https://orcid.org/0009-0002-2117-8951
e-mail : 20221080@sungshin.ac.kr
2022년 ~ 현 재 성신여자대학교 융합보안공학과 학사과정
관심분야 : Information Security

박 소 현

https://orcid.org/0009-0005-6925-9349
e-mail : 220227022@sungshin.ac.kr
2020년 성신여자대학교 융합보안학과(학사)
2022년 성신여자대학교 미래융합기술공학과 (석사)
2022년 ~ 현 재 성신여자대학교 미래융합기술공학과 박사과정
관심분야 : Wireless Communication Security, Digital Identity

이 일 구

https://orcid.org/0000-0002-5777-4029
e-mail : iglee@sungshin.ac.kr
2003년 서강대학교 전자공학과(학사)
2005년 한국과학기술원 정보통신공학과 (석사)
2016년 한국과학기술원 정보보호대학원 (박사)
2017년 ~ 현 재 성신여자대학교 융합보안공학과/ 미래융합기술공학과 부교수
관심분야 : Information Security, Wireless Networks and Communications