이 일 구 교수 지도
석사학위 청구논문

이산 웨이블릿 변환과 특이값 분해를 이용한 비가시성 워터마킹

2025

성신여자대학교 대학원
미래융합기술공학과
김 서 이

이산 웨이블릿 변환과 특이값 분해를 이용한 비가시성 워터마킹

이 일 구 교수 지도

이 논문을 석사학위논문으로 제출함

2024년 12월

성신여자대학교 대학원

미래융합기술공학과

김 서 이

인 준 서

김서이의 석사학위 논문으로 인준함

2024년 12월

심사위원장 김 성 민 (인)

심 사 위 원 임 연 섭 (인)

심 사 위 원 이 일 구 (인)

성신여자대학교 대학원

논 문 개 요

AI 기술의 발전과 함께 디지털 이미지의 저작권 침해 문제가 더욱 심각해지고 있으며, 이를 해결하기 위한 핵심 기술로 디지털 워터마킹 기술이 주목받고 있다. 디지털 워터마킹은 디지털 저작물의 저작권을 보호하거나 위·변조를 감별하고 추적하기 위해서 특수한 형태의 워터마크를 삽입하고, 검출하는 기술적 방법이다. 워터마크는 원본 이미지의 시각적 품질을 저하하지 않는 비가시성을 갖추어야 하며, 이미지 변형 및 노이즈 공격과 같은 다양한 외부 요인에 대해 강한 저항성을 유지하여 안정적으로 추출될 수 있어야 한다. 그러나 종래 연구에서는 비가시성을 강화하면 저항성이 저하되거나, 반대로 저항성을 강화하면 이미지 품질이 저하되는 트레이드오프 문제를 해결하지 못했다. 본 연구에서는 이미지에 3레벨 이산 웨이블릿 변환(Discrete Wavelet Transform, DWT)을 수행하여 주파수 성분을 여러 수준으로 분리한 후, 여러 영역에 걸쳐 특이값 분해 (Singular Value Decomposition, SVD)를 수행하여 특이값에 워터마크를 반복적으로 삽입하는 방식을 제안한다. 이를 통해 신호 변형 공격에 강인하고 비가시성이 뛰어난 워터마킹을 수행하는 것을 목표로 한다.

목 차

표 차 례

그림 차례

Ⅰ. 서 론

디지털 이미지의 무단 복제와 저작권 침해 문제는 디지털 미디어의 급격한 확산과 더불어 더욱 심각해지고 있다. 특히 생성형 AI 기술이 발전하면서 이미지를 변형하고 도용하는 사례가 급증하면서 디지털 컨텐츠의 저작권 침해 문제는 중요한 이슈로 부상하고 있다. 최근 이미지 제공 플랫폼과 여러 디지털 컨텐츠 아티스트들이 AI 기반 이미지 생성 플랫폼 서비스 기업에 대해 소송을 제기하는 사례가 늘어나고 있다[1]. 디지털 컨텐츠 저작권자들은 AI 학습에 자신들의 이미지를 사용하는 것이 저작권 침해에 해당하며, 적절한 라이선스 동의 없는 데이터 사용이 불법이므로 저작권을 보호해야 한다고 주장한다. 그러나 2022년에 미국 저작권청(U.S. Copyright Office)은 AI가 생성한 이미지는 저작권 보호 대상이 될 수 없다고 결정하였다[2]. 이 결정은 AI가 인간 창작자가 아닌 독립적인 주체로서 저작권 보호를 받을 수 없다는 점을 명확히 하였으며, 인간 창작자의 역할이 저작권의 핵심 원칙이라는 점을 재확인한 것이다. 이로 인해 AI로 생성된 이미지가 법적 보호를 받을 수 있는지에 대한 논란이 촉발되었고, 이는 AI 기술의 발전이 예술 및 창작물의 저작권 보호 방식에 큰 변화를 가져오고 있음을 시사한다. 동시에 이러한 변화는 디지털 이미지의 무결성과 저작권 보호를 위한 기술적 해결책의 필요성을 한층 더 부각시키고 있으며, 그중 하나로 디지털 이미지 워터마킹 기술이 주목받고 있다[3].

이와 같은 배경에서 디지털 이미지 워터마킹에 관한 연구가 활발히 진행되고 있다. 디지털 워터마킹은 저작권 보호 및 이미지 변조 감지와 같은 목적으로 이미지를 보호하는 효과적인 방법으로 주목받고 있으며, 이미지의 불법 복제 및 변조를 방지하는 중요한 도구로 널리 사용되고 있다[4].

디지털 워터마킹의 중요한 두 가지 요소는 원본 이미지의 품질을 훼손하지 않아야 하며, 워터마크가 외부 변형으로부터 강한 저항성을 가져야 한다[5]. 디지털 이미지 워터마킹은 이미지 품질을 유지하고 저항성을 갖기 위해 다양한 방식으로 구현될 수 있으며, 대표적으로 공간 도메인과 주파수 도메인을 활용하는 방법이 있다. 공간 도메인을 사용하는 워터마킹은 이미지의 픽셀값을 직접적으로 수정하여 워터마크를 삽입하는 방식으로, 상대적으로 구현이 간단하고 계산 비용이 낮아 실시간 처리에 적합하다는 장점이 있다[6]. 공간 도메인을 이용한 워터마킹은 이미지 변형에 취약하며, 압축이나 변조 시 워터마크가 손실될 가능성이 크다는 한계가 있다. 반면 주파수 도메인을 이용한 워터마킹은 이미지의 주파수 영역에 워터마크를 삽입하여 원본 이미지의 시각적 품질을 유지하면서도 높은 저항성을 제공하는 방식이다. 이미지를 주파수 도메인으로 변환하기 위해 이산 코사인 변환(Discrete Cosine Transform, DCT), 이산 웨이블릿 변환 (Discrete Wavelet Transform, DWT) 등의 방식이 이용되며 특정 성분에 대한 가공 및 처리를 수행할 수 있다. 그러나 이 방식 또한 특정 유형의 공격에 취약할 수 있다[7].

기존 방식의 한계를 극복하기 위해 다양한 이미지 처리 기법이 결합되어 활용되고 있다. DWT와 같은 주파수 변환 기법은 주파수 대역을 분리하여 워터마크 삽입에 유리한 위치를 제공하지만, 특정 공격에 약점을 가질 수 있다. 이를 보완하기 위해 최근에는 주파수 변환 기법과 더불어 특이값 분해(Singular Value Decomposition, SVD) 기법 등이 함께 사용되고 있다[8]. SVD는 이미지의 주요 정보가 집중된 특이값을 조정해 워터마크를 삽입하는 방식으로, 이미지 품질을 유지하면서도 워터마크의 강인성을 크게 향상시킬 수 있는 장점이 있다[9]. DWT-SVD 결합 방식은 DWT를 통해 이미지를 저주파와 고주파 대역으로 나누고, 특정 대역의 특이값에 워터마

크를 삽입함으로써 노이즈나 압축 공격에 대한 저항력을 강화할 수 있다. 이러한 결합 방식은 다양한 공격에도 워터마크가 견딜 수 있도록 해주어 최근 연구에서 널리 사용되고 있다[10],[11].

본 연구에서는 이미지에 3 Level DWT를 수행하여 주파수 성분을 여러 수준으로 분리한 후, 저주파 영역 및 일부 고주파 영역에 특이값 분해를 적용하여 저주파 성분의 특이값에 워터마크를 반복적으로 삽입한다. 저주파 영역은 이미지의 구조적 정보를 담고 있어서 약간의 변화로도 이미지 전체에 큰 영향을 줄 수 있지만, 미세한 변화를 가함으로써 비가시성을 유지하면서 고주파 영역에서 발생할 수 있는 문제를 회피할 수 있다. 또한 일부 고주파 영역의 특이값에도 워터마크를 반복적으로 삽입함으로써 특정 주파수에 대한 필터링 공격 등에 대한 저항성을 확보할 수 있다. 추출 과정에서는 에러 정정 기능을 포함하여 발생할 수 있는 오류를 보정하며, 반복적으로 임베딩된 워터마크의 상관성에 기반하여 워터마크를 탐지한다.

본 연구의 주요 기여점은 다음과 같다.

1) 이미지 저작권 보호를 위한 3레벨 DWT 및 SVD를 결합한 비가시성 워터마킹 방식을 제안한다.

2) 여러 주파수 성분에 워터마크를 반복적으로 삽입하여 일부가 훼손되어 오류가 발생하더라도 정정할 수 있도록 한다.

3) 다양한 신호 변형 기법에 대한 강인성을 갖는다.

4) 종래 방식 대비 이미지 품질 및 워터마크 추출 성능을 개선한다.

본 논문의 구성은 다음과 같다. Ⅱ절에서는 DWT, SVD 및 신호 변형 기법 등 본 논문의 배경 지식을 설명하고, Ⅲ절에서는 관련 연구를 분석한다. Ⅳ절에서는 3레벨 DWT 및 SVD를 결합한 비가시성 워터마킹 방식을 제안하며, Ⅴ절에서는 실험 환경 및 과정을 설명한다. Ⅵ절에서는 성능 평가를 위한 실험 결과를 분석한다. 마지막으로 Ⅶ절에서 결론을 맺는다.

Ⅱ. 배경 지식

2.1. 이산 웨이블릿 변환 (Discrete Wavelet Transform, DWT)

이미지 처리 시 고려할 수 있는 도메인은 공간 도메인과 주파수 도메인이다. 공간 도메인은 이미지를 구성하는 픽셀값에 직접 접근하여 처리하는 영역으로, 각 픽셀의 밝기나 색상 정보를 기반으로 이미지를 처리한다. 밝기 조절, 대비 조정 및 공간 필터링과 같은 연산 과정은 공간 도메인에서 이루어지며, 이미지의 물리적 특성을 그대로 반영한다[12]. 반면 주파수 도메인은 이미지를 구성하는 변화를 주파수 성분으로 변환하여 분석하고 처리하는 영역이다. 이미지를 저주파 성분과 고주파 성분으로 분리하여 처리할 수 있기 때문에 노이즈 제거 및 압축 등의 작업을 수행하기에 유리하다[13].

이미지를 처리할 때 초기에는 공간 도메인에서 작업이 이루어지며, 주파수 성분을 분석하거나 조작하려면 주파수 도메인으로 변환하는 과정이 필요하다. DWT는 이미지를 주파수 도메인으로 변환하는 기법 중 하나로 주파수 성분과 공간적 위치 정보를 동시에 분석할 수 있는 특징이 있다[14]. 즉, 이미지를 주파수 성분으로 변환하면서도 공간적인 위치를 유지해 이미지의 다양한 변화를 여러 해상도에서 분석할 수 있는 이점이 있다[15]. Figure 1과 Figure 2는 이미지에 DWT를 적용하여 주파수 영역으로 변환하는 방법이다.

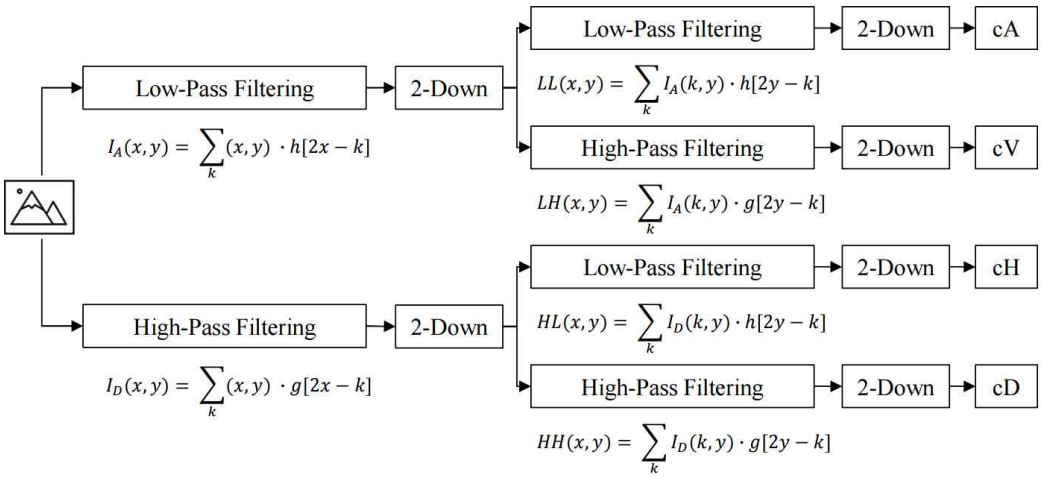


Figure 1. 3-Level Discrete Wavelet Transform

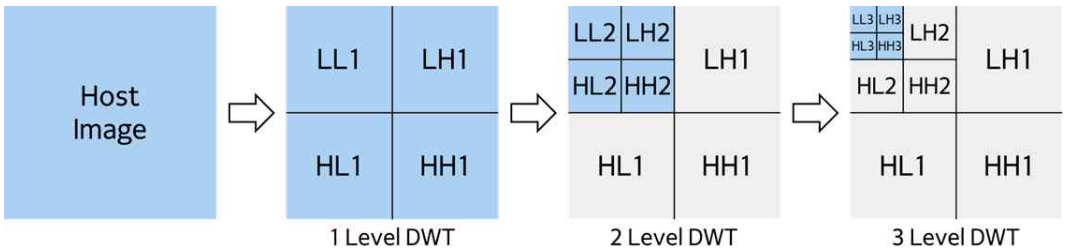


Figure 2. Discrete Wavelet Transform Diagram

이미지를 공간 도메인에서 주파수 도메인으로 변환할 때, DWT 과정은 필터링과 다운 샘플링을 통해 수행된다. 1차원 DWT는 저주파 필터(Scaling Function)와 고주파 필터(Wavelet Function)를 사용하여 주어진 신호를 저주파 성분(Approximation Coefficient)과 고주파 성분(Detail Coefficient)으로 변환한다. 이미지의 경우 2차원 DWT가 수행되는데, 이미지의 각 행과 열에 1차원 DWT를 2번 수행한다. 먼저 행 방향으로 DWT가 적용되어 저주파 성분과 고주파 성분으로 구분되고, 여기에 열 방향으로 DWT가 적용되어 저주파-저주파(LL), 저주파-고주파(LH), 고주파-저주파(HL), 고주파-고주파

(HH)로 구분되어 4개의 서브밴드가 생성되게 된다. LL은 근사 계수(Approximation Coefficients, cA)로 이미지의 전반적인 특성을 나타낸다. LH, HL은 각각 수평 세부 계수 (Horizontal Detail Coefficients, cH), 수직 세부 계수(Vertical Detail Coefficients, cV)에 해당되며, 세로 방향의 경계 정보와 가로 방향의 경계 정보를 담고 있다. HH는 대각 세부 계수(Diagonal Detail Coefficients, cD)로 대각선 방향에서의 경계나 세부적인 정보를 담고 있다. cA를 저주파 영역으로, 나머지 cH, cV, cD를 고주파 영역으로 분류한다.

2.2. 특이값 분해 (Singular Value Decomposition, SVD)

SVD는 이미지의 정보가 분산되는 방식을 파악하고 이를 활용할 수 있도록 하여 DWT 등의 변형 기법과 함께 디지털 이미지 워터마킹 기법에 활발히 이용된다[16], [17]. SVD는 이미지 행렬을 세 개의 행렬로 분해하는데, 직교 행렬 U, 대각 행렬 S, 직교 행렬 V로 구성된다.

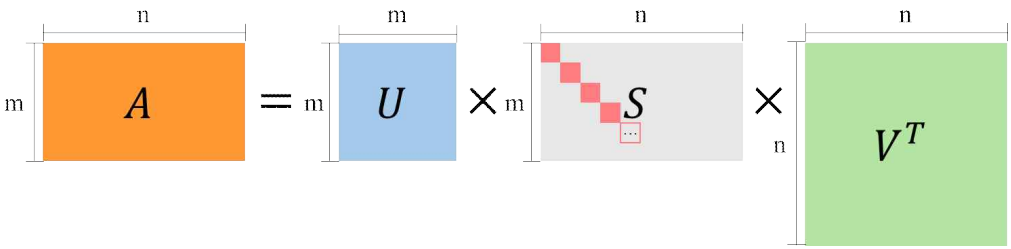


Figure 3. Singular Value Decomposition

직교 행렬 U는 이미지(A)의 왼쪽 특이 벡터로 구성되며, V는 오른쪽 특이 벡터로 구성된다. 대각 행렬 S의 대각선 원소에 해당되는 값이 특이값

이며, 행렬의 크기 및 정보 등 이미지의 주된 정보를 나타낸다. 작은 특이값을 제거할수록 정보 손실이 발생할 수 있지만 데이터의 주요 구조는 유지되기 때문에 신호 처리 및 이미지 처리 등의 분야에서 데이터 구조를 유지한 채 노이즈를 줄이거나 압축하기 위해 이용한다[18].

워터마킹 분야에서 SVD는 특이값을 변형하여 워터마크를 삽입하는 형태로 이용된다. 호스트 이미지의 특이값과 워터마크 이미지의 특이값을 결합하거나 조정하는 방식으로 최종 워터마크 이미지의 특이값을 구성한다. 특이값에 작은 변형을 가하면 시각적으로 인식하기 어렵기 때문에 호스트 이미지의 품질을 유지하며 워터마크를 삽입할 수 있다[19]. 또한 이미지의 중요한 정보는 특이값에 압축된 형태로 표현되기 때문에, 외부 공격으로 인한 이미지 변형이 발생했을 때 워터마크가 그대로 유지될 가능성이 높다. 특이값이 이미지의 고유 특성을 반영하고 왜곡이나 공격에 강하므로 워터마크 삽입 시 큰 이점을 제공한다[20].

DWT와 SVD를 결합하면 디지털 이미지 워터마킹 시 워터마크 강인성 및 이미지 품질을 개선할 수 있다. DWT를 통해 이미지의 주파수 성분을 여러 수준으로 분리하여 저주파(LL)와 고주파(LH, HL, HH) 성분으로 나눌 수 있으며, 저주파 성분은 이미지의 주요 정보를 포함하고 있다[21]. 이로 인해 외부 공격에 대해 더 견고하며, 저주파 성분에 워터마크를 삽입하게 되면 압축이나 필터링, 노이즈 등의 공격에 대해 강인성을 향상시킬 수 있다. SVD를 추가로 적용하면 저주파 성분의 특이값에 워터마크를 삽입하게 되어 저주파 성분의 주요 정보에 접근할 수 있다. 특이값은 작은 변형이 가해져도 비교적 안정적인 구조를 유지할 수 있기 때문에 DWT와 SVD를 결합하면 저주파 대역의 강인성 및 외부 저항성을 극대화할 수 있다.

2.3. 신호 변형 기법 (Signal Distortion Techniques)

디지털 워터마킹 알고리즘을 설계할 때, 신호 변형 공격에 대한 저항성을 고려하는 것이 중요하다. 실제로 이미지에 대한 불법적인 접근, 이용, 가공 등을 목적으로 노이즈 공격을 수행하여 디지털 워터마크를 무력화하는 시도가 빈번하게 발생하고 있다. 따라서 디지털 워터마킹 알고리즘을 설계하는 과정에서 외부 공격에도 워터마크 정보가 유지되도록 강인성을 갖추어야 한다[22]. 대표적인 신호 변형 기법에는 노이즈 공격, 압축 공격, 필터링 공격 등이 있다.

1) 노이즈 공격 (Noise Attack)

노이즈 공격은 워터마크가 삽입된 이미지에 랜덤한 노이즈를 추가하는 방식의 공격 기법이다. 노이즈는 이미지의 각 픽셀값에 영향을 미치기 때문에 워터마크의 가시성 및 강인성에 영향을 미친다[23].

가우시안 노이즈(Gaussian Noise)는 가장 일반적인 랜덤 노이즈 공격이며 노이즈가 정규 분포 형태를 따른다. 주로 이미징 센서의 열 소음이나 낮은 조명 등 이미지 처리 과정에서 발생할 수 있는 물리적인 환경으로 인한 노이즈를 모방한다. 모든 픽셀에 동일한 크기의 노이즈가 랜덤하게 추가되어 이미지 전반에 퍼지게 되며, 고주파 영역의 정보 품질을 저하시킨다[24].

소금&후추 노이즈 (Salt & Pepper Noise)는 이미지에 검은색(0)과 흰색(255) 픽셀을 무작위로 추가하는 방식으로, 이미지에 극단적인 노이즈가 추가된다. 이 공격 방식은 이미지의 선명도와 대비에 큰 영향을 미치기 때문에 이미지의 명암 대비가 강조되며, 원본 정보와 워터마크의 명확성을 저

하시킨다. 워터마크가 특정 픽셀의 밝기 변화에 기반하면 무결성이 심각하게 훼손될 수 있으며, 이미지 경계 및 패턴에 워터마크가 삽입되었을 때도 취약하다[25].

스파클 노이즈(Sparkle Noise)는 임의의 밝은 점이 이미지에 무작위로 추가되는 방식으로, 주로 이미지의 특정 영역에 밝은 픽셀이 집중적으로 발생한다[26]. 이미지의 픽셀 값에 스케일링된 가우시안 노이즈를 곱하는 방식으로 공격이 수행되며 원본 정보의 형태가 왜곡되며 이미지가 불균일해진다. 워터마크가 이미지의 전체적인 구조나 텍스처에 영향을 받는 경우 스파클 노이즈에 의해 쉽게 훼손될 수 있으며, 이미지 텍스처가 중요한 고해상도 워터마킹 방식이나 SVD 기반 워터마킹 기법에서도 강한 공격 효과를 보인다.

2) 압축 공격 (Compression Attack)

압축 공격은 이미지의 용량을 줄이는 과정에서 발생하는 손실을 통해 워터마크의 강인성을 평가하는 방법이다. 손실 압축 방식의 경우 이미지의 일부 정보를 탈락시키는 방식으로 압축을 수행하기 때문에 이 과정에서 워터마크 정보가 손실될 수 있다. 또한 이미지 전송 및 가공 과정에서 압축이 수행되며 악의성 없이도 압축으로 인한 워터마크 손실이 발생할 수 있기 때문에 워터마크 설계 시 중요하게 고려되어야 한다.

이미지 압축 표준인 JPEG는 이미지의 주파수 성분을 DCT로 변환하고, 고주파 성분을 제거하는 방식으로 파일 크기를 줄이는 손실 압축 방식이다[27]. 고주파 영역의 정보를 많이 제거하며 압축률이 높을수록 저주파 정보도 영향을 받을 수 있다.

JPEG2000은 DWT 변환을 기반으로 한 압축 방식으로 손실 및 무손실

압축을 모두 지원하기 때문에 더 높은 품질의 이미지를 제공할 수 있다. 공격 수행 시에는 주로 손실 압축 방식이 이용되며 주파수 성분의 일부를 제거한다[28]. DWT 기반 압축은 저주파 성분에도 영향을 주기 때문에 JPEG2000 손실압축 방식을 이용하면 저주파 영역에 삽입된 워터마크가 훼손될 수 있다.

3) 필터링 공격 (Filtering Attack)

필터링 공격은 이미지에 특정 필터를 적용하여 워터마크의 강인성을 평가하는 방식으로 공격 과정에서 이미지의 세부 정보나 특정 주파수의 성분이 제거된다.

블러링 공격은 이미지에 블러 필터를 적용하여 모서리나 세부 텍스처가 흐려지도록 유도하여 워터마크의 신호를 약화시키거나 제거하는 방식의 공격이다. 블러링은 이미지의 세부 정보를 담고 있는 고주파 성분을 주로 제거하기 때문에 워터마크가 고주파 대역에 삽입된 경우, 쉽게 손상될 수 있다[29].

저주파 필터링 공격은 이미지의 고주파 성분을 제거하여 이미지의 부드러움을 유지하면서 세부 사항을 줄이는 필터링 방식이다. 저주파 필터는 노이즈 제거에 이용되며, 이미지를 평활하게 하여 경계선이나 디테일 등 고주파 요소를 제거한다[30]. 워터마크가 고주파 영역에 삽입된 경우, 저주파 필터링에 의해 쉽게 손상될 수 있다.

Ⅲ. 관련 연구

디지털 멀티미디어의 저작권에 관한 관심이 높아지면서 저작물에 대해 효과적인 보호 방안을 찾기 위한 노력이 계속되고 있다. 그중 하나로 주목받고 있는 디지털 워터마킹과 관련해서도 연구가 활발히 수행되며 새로운 워터마킹 기법이 제안되고 있다. 본 장에서는 주파수 도메인에서 비가시성 및 강인성을 갖춘 다양한 워터마킹 기법을 조사하고 분석한다.

1) DWT를 이용한 디지털 이미지 워터마킹 기법

Hosseini 외 2인의 연구[31]에서는 외부 공격에 강인한 디지털 이미지 워터마킹을 위해 다양한 주파수 변환 방식을 결합한 새로운 하이브리드 워터마킹 방식을 제안했다. DWT, DCT, PCA를 결합한 방식은 다중 해상도 분석과 데이터 압축 기법을 활용하여 노이즈, 압축 등의 공격에 강인성을 개선했다. 하지만 종래 연구와 비교했을 때 높은 계산 복잡도 및 자원이 요구되며, 강인성과 비가시성 사이의 균형을 이루어 내지 못해서 호스트 이미지가 크게 훼손되는 한계가 있다.

Lidyawati 외 4인의 연구[32]에서는 호스트 이미지에 3레벨 DWT를 수행한 후, 저주파 영역(LL3)에 워터마크를 임베딩하는 워터마킹 방식을 제안했다. 워터마킹 임베딩 시, 임베딩 강도를 크게 낮추어 호스트 이미지의 품질을 우수하게 유지했다. 워터마크의 강인성을 평가하기 위해 가우시안 노이즈, Salt & Pepper, 블러링 공격을 수행했지만, 공격 강도를 고려하지 않았고, 일부 공격에서는 워터마크 복원 성능이 우수하지 못했다. 해당 논문의 제안 방식은 워터마크의 투명성을 강조했으나, 외부 변형에 대한 강인성을 갖추었다고 평가하기엔 한계가 있다.

2) DWT 및 SVD를 결합한 디지털 이미지 워터마킹 기법

Kusumaningrum 외 3인의 연구[33]에서는 2레벨 DWT와 SVD를 결합한 이미지 워터마킹 방식을 제안했다. 저주파 영역(LL2)에 워터마크를 임베딩했으며, 추출 시엔 논블라인드 추출 방식을 이용한다. DWT 및 SVD를 단독으로 이용한 경우와 제안 방식을 비교하는 실험을 수행하였고, Salt&pepper, gaussian filter, JPEG 압축 공격 등을 이용하여 강인성 평가를 진행했다. 본 논문에서는 공격 수행 시 공격 강도에 대한 고려가 이루어지지 않았고, 제안한 워터마킹 방식의 강인성을 입증할만한 공격이 고려되지 않았다. 또한 제안 방식은 비교된 두 단일 모델에 비해 향상된 워터마킹 추출 성능을 보였지만, 일부 공격을 수행했을 때 비교적 낮은 추출 성능을 보이는 한계점이 있다.

Yasmeen 외 1인의 연구[16]에서는 저작권 보호를 위한 비가시성 및 강인성을 갖춘 워터마킹 방식을 제안했다. 호스트 이미지에는 4레벨 DWT를 수행하고 저주파(LL4)와 고주파(HH4) 성분에 SVD를 수행했으며, 워터마크 이미지에는 3레벨 DWT를 수행한 후 저주파(LL3)와 고주파(HH3) 성분에 SVD를 수행했다. 호스트 이미지의 LL4, HH4의 특이값과 워터마크 이미지의 LL3, HH3 특이값을 결합하는 형태로 워터마크를 임베딩했다. 워터마킹 삽입 후, 호스트 이미지의 품질은 평균 40dB를 유지했다. 강인성 평가를 위해 다양한 공격을 수행했으며, PSNR 측면에서 가우시안 노이즈 공격은 평균 23dB, Salt & Pepper 노이즈 공격은 평균 27.5dB, 스파클 노이즈 측면에서 평균 30dB의 추출 성능을 보였다. 비가시성 및 강인성 측면에서 전반적으로 우수한 성능을 보였지만 수행된 공격의 강도를 알 수 없으며, 다양한 강도의 공격에 의한 영향에 대한 분석이 부족했다.

Kodathala 외 2인의 연구[34]에서는 디지털 미디어의 저작권 보호를 목

적으로 외부 변형으로부터 강인한 워터마킹 방식을 제안한다. 호스트 이미지와 워터마크 이미지에 2레벨 DWT를 수행한 후 저주파 영역(LL2)에 SVD를 수행하여 호스트 이미지 특이값에 워터마크 이미지의 특이값을 결합하는 방식으로 워터마크 임베딩을 수행한다. 강인성 평가를 위해 다양한 공격이 고려되었으며, 여러 최근 연구와의 비교를 수행했다. 전반적으로 기존 연구 대비 향상된 추출 성능을 보였지만 가우시안 블러, 스파클 노이즈 등 일부 공격이 수행된 후 추출된 워터마크에서 복원성의 한계를 보였다.

Araghi 외 1인의 연구[17]에서는 SVD 레벨에 따른 디지털 워터마킹 기법의 효과를 분석하고 새로운 워터마킹 방식을 제안했다. 호스트 이미지에 1레벨 DWT를 수행한 후 전체 주파수 영역(LL1, LH1, HL1, HH1)에 SVD를 수행하여 워터마크 이미지를 결합하는 방식의 워터마킹 기법을 제안했으며, SVD 수행 시, 1레벨 SVD 방식과 2레벨 SVD 방식을 각각 적용하여 비교 분석하는 과정을 포함하고 있다. 2레벨 SVD를 수행한 경우, 1레벨 SVD 방식 대비 호스트 이미지 품질이 매우 우수하게 유지되었으나 일부 공격에서 워터마크가 잘 추출되지 않는 한계가 있었다.

Ⅳ. DWT 및 SVD를 결합한 이미지 워터마킹 기법

본 논문에서는 신호 변형 기법에 강인한 비가시성 워터마킹 기법을 제안한다. 제안한 방식은 3레벨 이산 웨이블릿 변환을 통해 이미지를 주파수 성분을 여러 수준으로 분리한 후, 이미지의 저주파 영역 및 일부 고주파 영역에 SVD를 적용한다. 특이값에 워터마크를 반복적으로 삽입하여 신호 변형 기법을 이용한 이미지 공격으로 인한 워터마크의 훼손을 방지한다. 여러 주파수 구역에 삽입된 워터마크를 추출하고 이를 효과적으로 통합하여 워터마크를 성공적으로 복원한다.

이미지의 고주파 영역은 이미지의 경계를 표현하는 세부 정보를 담고 있어서 이 영역이 변형되면 이미지가 왜곡되거나 품질에 영향을 미칠 수 있다. 그러나 고주파 성분에 약간의 왜곡이나 변형이 발생해도 인간의 시각으로는 구분하기 어려운 특성이 있다. 이러한 특성을 활용하여 최근에는 고주파 영역을 활용한 비가시성 워터마크 기술에 관한 연구가 진행되고 있으며, 이러한 워터마크 기법이 적용된 디지털 컨텐츠의 고주파 성분을 타겟으로 한 노이즈 공격도 연구되고 있다. 공격자는 고주파 영역을 이용해 시각적으로 눈에 띄지 않게 워터마크를 방해할 수 있는 것으로 보고되었다.

반면, 저주파 영역은 이미지의 전체적인 형태와 주요 객체의 중요한 정보를 담고 있어서 이 영역이 변형되면 이미지 구조에 큰 변화를 초래할 수 있다. 따라서 저주파 영역에 워터마크를 삽입할 때는, 미세한 변형을 통해 이미지의 품질 저하를 최소화하는 것이 중요하다. JPEG 압축을 활용한 공격은 고주파 성분을 주로 제거하므로 저주파 영역에 임베딩된 워터마크는 고주파 영역에 임베딩된 워터마크보다 압축 기반 공격 기법에 상대적으로 강인하며, 가우시안 노이즈나 저해상도 변환과 같은 노이즈 공격에 덜 민

감하다. 또한, 이미지 변환 후에도 저주파 성분은 중요한 정보를 유지하므로, 저주파 영역에 워터마크를 삽입할 경우, 큰 이미지 변형이 발생하지 않는 한 다양한 공격에도 워터마크를 복구할 수 있다. 하지만 JPEG의 압축률이 높아질수록 저주파 성분도 일부 손실될 수 있으며, JPEG2000과 같이 저주파 대역까지 압축하는 방식에서는 저주파 성분에 삽입된 워터마크가 더 큰 영향을 받을 수 있다. 이미지의 밝기나 대비와 같은 전반적인 특성에 대한 변화가 발생하는 경우에도 저주파 영역의 워터마크에 영향을 미친다.

본 연구에서는 이러한 특성을 활용하여 저주파 성분과 일부 고주파 성분을 동시에 이용하여 노이즈 공격에 강인한 비가시적 워터마킹 기법을 구현하였다.

4.1 워터마크 임베딩 과정

DWT와 SVD를 결합한 이미지 워터마킹 방식은 일반적인 구조는 유사하지만 연구의 목적과 방향에 따라 세부적인 과정에서 차이가 있다. 일반적으로 호스트 이미지에 DWT를 적용하여 서브밴드(LL, LH, HL, HH)를 생성한 후, 특정 서브밴드에 대해 SVD를 수행한다. 워터마크는 SVD에서 얻은 특이값을 수정하는 방식으로 삽입되며, 이 과정에서도 각 연구마다 방법론적 변형이 나타날 수 있다. 강건성과 비가시성의 강화, 계산 효율성의 향상 등 다양한 목표를 달성하기 위해 다양한 방법이 이용될 수 있다.

Figure 4는 제안한 워터마크 삽입 과정이다. 제안된 방식에서는 호스트 이미지의 크기에 따라 고정된 워터마크 이미지의 크기가 요구된다. 호스트 이미지 크기에 따른 워터마크 이미지의 크기는 수식(1)과 같이 정의한다.

$$W = \frac{N}{2^L} \tag{1}$$

W는 워터마크 이미지의 한 변의 크기이고, N은 호스트 이미지의 한 변의 크기에 해당된다. L은 DWT의 레벨을 의미한다. 본 논문에서는 호스트 이미지의 한 변의 크기가 512이고, 3레벨 DWT를 수행했으므로 64x64 픽셀 크기의 워터마크 이미지를 이용해야 한다.

호스트 이미지에 3레벨 DWT를 수행하여 주파수 영역으로 변환하면 LL3, LH3, HL3, HH3 서브밴드를 생성한다. 이 중 저주파 영역(LL3)과 일부 고주파 영역(LH3, HL3)에 SVD 특이값 분해를 수행한다. LL3, LH3, HL3에서 각각 얻어진 특이값에 워터마크를 임베딩한 후, 이 수정된 특이값(St)에 대해 다시 한번 SVD를 수행한다. 이렇게 얻어진 특이값(Sw)을 이용하여 LL3(LL3t), LH3(LH3t), HL3(HL3t)를 재구성하고 3레벨 이산 웨이블릿 역변환(Inverse Discrete Wavelet Transfer, IDWT)을 통해 워터마

크가 삽입된 이미지를 구성한다. 워터마크가 삽입된 이미지는 동일한 워터마크가 각각 다른 주파수 성분에 3번 반복된 형태로 구성된다.

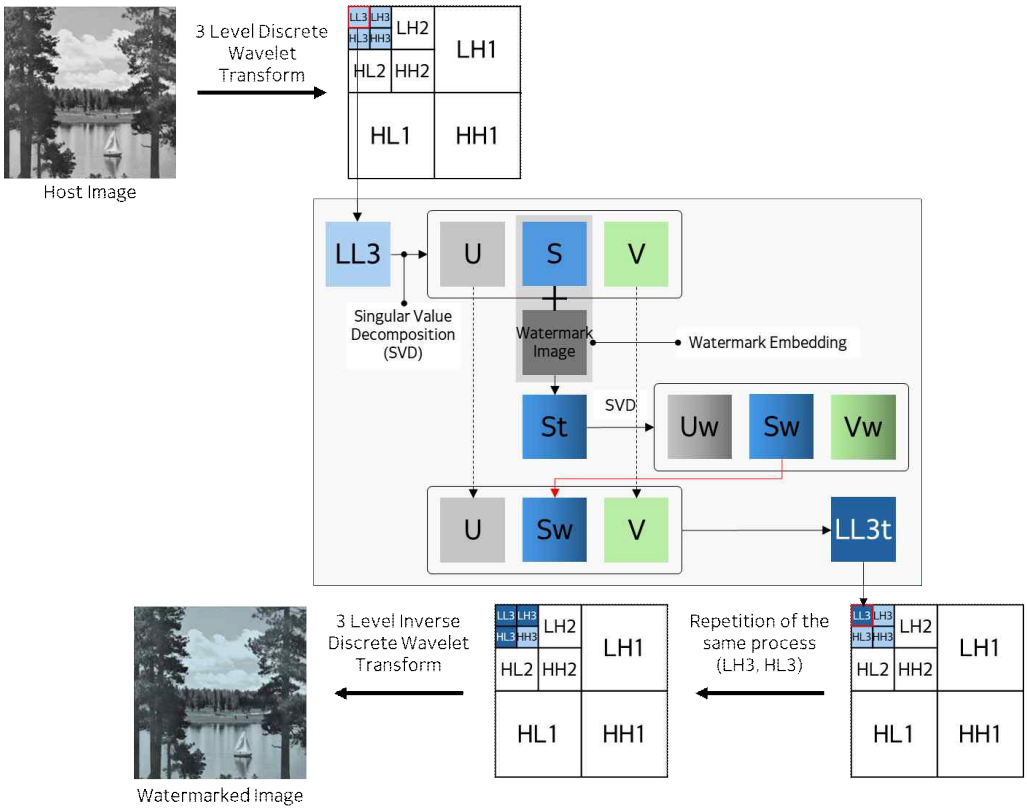


Figure 4. Watermark Embedding Process

본 연구에서 제안한 방법은 호스트 이미지에 DWT를 수행한 후, SVD를 적용하여 특이값에 워터마크 이미지를 임베딩하고, 다시 SVD를 수행하여 갱신된 특이값을 활용한다. 워터마크 임베딩 과정에서 특이값이 수정되면서 본래 호스트 이미지의 구조적 특징에 일부 변화가 발생할 수 있다. 그

러나 수정된 특이값이 이미지의 구조와 조화를 이루지 못할 경우, 이미지 품질이 저하되거나 워터마크 복원이 어려울 가능성이 존재한다. 이러한 문제를 해결하기 위해 두 번째 SVD 과정을 추가로 수행하며, 이는 수정된 특이값이 원래 이미지의 구조에 자연스럽게 융합되도록 돕는다. 이 재조정 과정(re-calibration)을 통해 특이값은 조화롭게 재배치되어 이미지의 원래 구조에 깊이 통합되며, 결과적으로 워터마크의 비가시성과 강인성을 향상시킨다.

4.2 워터마크 추출 과정

Figure 5는 제안한 워터마크 추출 과정이다. 추출 과정은 논블라인드 워터마킹 방식으로 수행된다. 워터마크가 임베딩된 이미지에 3레벨 DWT를 수행하여 주파수 영역으로 변환한다. LL3, LH3, HL3 성분에 각각 SVD 특이값 변환을 수행하여 워터마크가 삽입된 특이값 행렬(Sw)을 추출한다. 이렇게 추출된 각 성분의 특이값 행렬(Sw)과 Uw, Vw을 이용하여 워터마크 이미지를 재구성한다. LL3, LH3, HL3 성분에 워터마크가 한 번씩 임베딩 되었으므로 총 3개의 워터마크를 추출할 수 있으며, 이를 이용하여 최종 워터마크 이미지를 복원한다.

워터마크 복원은 다음과 같은 과정을 통해 수행된다. LH3와 HL3에서 추출된 워터마크 배열에서 동일한 위치의 값들에 대해 중간값 계산(Median Fusion)을 수행한다. 이 과정을 통해 LH3와 HL3 대역에서 얻은 정보를 결합하여 노이즈를 감소시키고 워터마크 데이터를 통합한다. 이렇게 얻어진 중간 워터마크 데이터와 LL3 대역에서 추출된 워터마크를 가중 결합(Weighted Combination)을 통해 결합한다. LL3 대역은 가장 중요한 정보를 포함하고 있고, 외부 변형이나 노이즈에 의한 영향을 가장 적게 받기 때문에 높은 가중치를 부여한다. 이 과정을 통해 LL3 대역의 정보를 중심적으로 활용하면서 LH3, HL3 대역에서 얻은 정보를 보완적으로 활용한다.

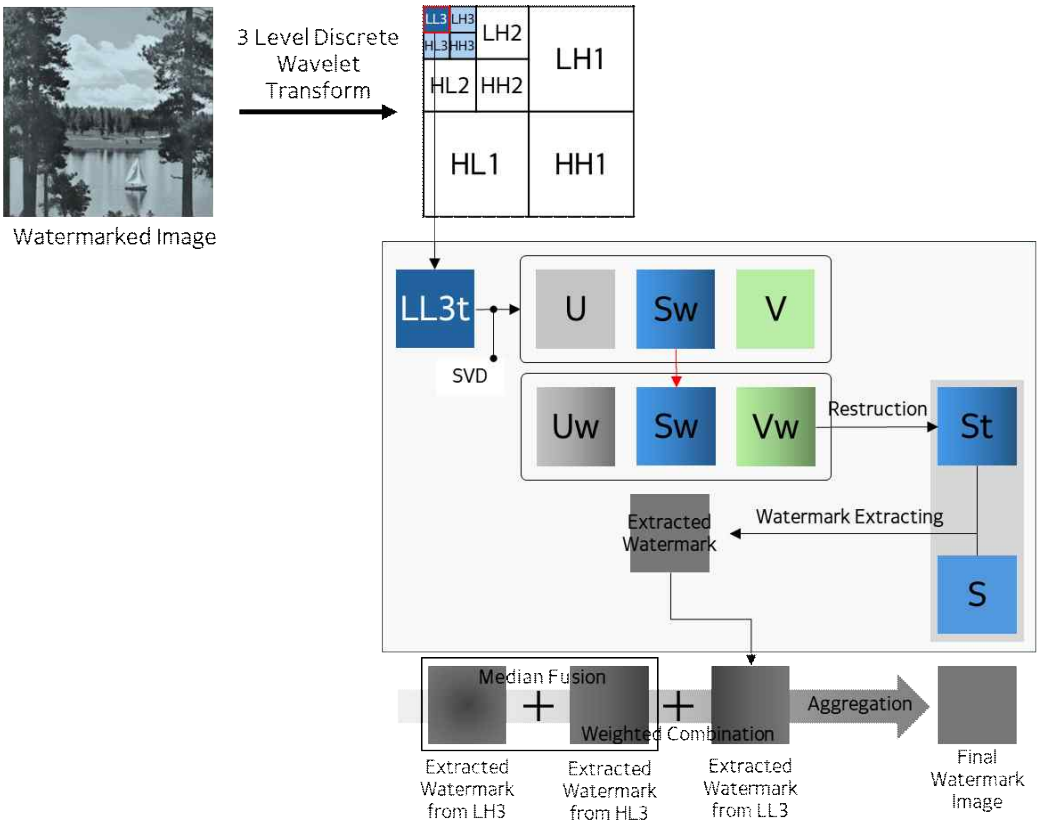


Figure 5. Watermark Extraction Process

Ⅴ. 실험 환경

제안한 3레벨 DWT 및 SVD를 결합한 이미지 워터마킹 기법의 성능 평가를 위한 실험을 수행하였다. 본 장에서는 실험 환경 및 과정, 이용된 성능 평가지표를 설명한다.

5.1. 실험 환경

Figure 6은 실험에 이용된 호스트 이미지와 워터마크 이미지이다. 호스트 이미지로 512x512 픽셀 크기의 그레이 스케일 이미지를 이용했으며 워터마크 이미지는 64x64 픽셀 크기의 그레이 스케일 이미지를 이용했다.

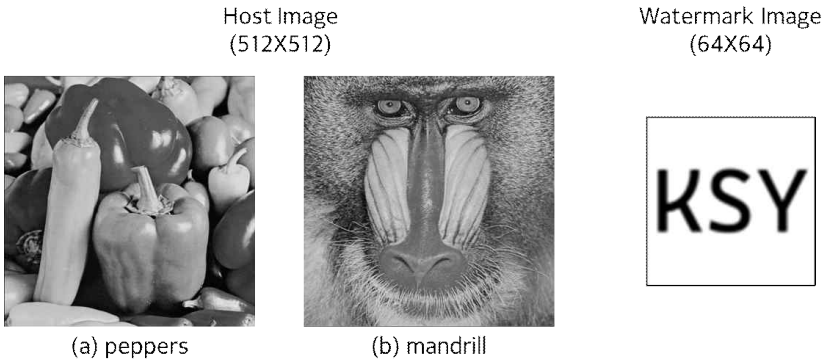


Figure 6. Host Image and Watermark Image

제안된 워터마킹 기법의 강인성을 평가하기 위해 총 7개의 신호 변형 공

격 기법을 이용하여 실험을 수행하였다. 노이즈 공격, 압축 공격, 필터링 공격 등을 수행했으며, 약한 수준의 공격부터 매우 강한 수준의 공격까지 5단계로 조정하며 강인성을 평가했다. 각 공격의 공격 강도 설정에 이용된 파라미터와 설정값은 Table 1과 같다.

Table 1. Attack Parameters and Attack Intensity

Attack type	Attack	Parameter	Attack intensity (level)				
			1 weak	2	3	4	5 strong
Noise Attack	Gaussian noise	Variance	0.001	0.005	0.01	0.05	0.1
	Salt and pepper	Density	0.01	0.03	0.05	0.1	0.2
	Sparkle noise	Probability	0.01	0.03	0.05	0.1	0.2
Compression Attack	JPEG compression	Quality factor	90	70	50	30	10
	JPEG2000 compression	Quality factor	90	70	50	30	10
Filtering Attack	Blurring attack	Kernel size	3	5	7	9	11
	Low-frequency Filtering	Kernel size	3	5	7	9	11

가우시안 노이즈 공격은 분산(Variance) 값을 통해 공격 강도를 설정했다. 분산을 통해 노이즈의 세기를 설정할 수 있으며, 값이 클수록 노이즈가 강해진다.

소금 & 후추 공격은 밀도(Density) 파라미터를 통해 공격 강도를 설정했는데, 이는 노이즈가 이미지에 영향을 미치는 픽셀의 비율을 의미한다. Density가 0.1일 경우, 소금(흰색, 255) 픽셀이 5%, 후추(검은색, 0) 픽셀이 5%로 총 10%의 픽셀에 노이즈가 추가된다.

스파클 노이즈 공격은 확률(Probability) 파라미터를 통해 특정 픽셀에

노이즈가 추가될 확률을 설정했다. 파라미터 값이 높을수록 더 많은 픽셀에 노이즈가 추가된다.

JPEG 압축 공격과 JPEG2000 압축 공격에서는 품질 지수(Quality Factor)를 이용하여 공격 강도를 설정했다. 품질 지수는 압축 후 이미지 품질을 결정하는 파라미터로, 값이 낮을수록 강한 압축을 수행하며, 이미지 품질이 저하된다.

블러링 공격과 저주파 필터링 공격에는 커널 크기(Kernel size)를 통해 공격 강도를 설정했다. 블러링 공격에서는 블러 필터의 크기를 의미하며, 필터가 적용되는 픽셀 영역의 크기가 된다. 커널 크기가 클수록 블러 효과가 강해지고, 이미지 정보에 더 큰 손실이 발생하게 된다. 저주파 필터링 공격에서는 저주파 필터의 범위를 결정하는 역할을 하며, 커널 크기가 클수록 더 많은 고주파 성분이 제거된다. 커널 크기가 3인 경우, 3X3 크기의 필터를 이용하게 된다.

각 공격 별로 적절한 파라미터를 이용하여 공격 강도를 조절했으며, 1단계의 약한 수준의 공격부터 5단계의 매우 강한 수준의 공격까지 적용하여 강인성 평가를 진행했다.

5.2. 실험 과정

Figure 7은 실험 과정에 대한 흐름도이다. 그레이 스케일의 호스트 이미지를 업로드한 후, 512x512에 맞추어 크기를 조정한다. 주파수 도메인으로 변환하기 위해 3레벨 DWT를 수행하여 LL3, LH3, HL3, HH3 성분에 접근할 수 있도록 한다. DWT는 Daubechies4 웨이블릿(db4)과 경계 처리 모드(periodization)를 사용하여 수행되었다. 이 중 저주파 영역인 LL3와 일부 고주파 영역(LH3, HL3)에 SVD 특이값 분해를 수행한다. 이를 통해 얻은 특이값 행렬을 이용하여 워터마크(W)를 임베딩 한다. 워터마크 임베딩 시에는 alpha(α) 값을 통해 워터마크 임베딩 강도를 설정할 수 있다. 위 과정을 통해 얻은 St 행렬에 다시 SVD 특이값 분해를 수행하여 워터마크가 임베딩된 특이값 행렬 Sw를 얻을 수 있다. $[U, Sw, V]$를 이용하여 워터마크가 삽입된 LL3t, LH3t, HL3t를 재구성한다. 기존의 LL3, LH3, HL3 성분 대신 LL3t, LH3t, HL3t를 이용하여 3레벨 IDWT를 수행하면 워터마크가 임베딩된 이미지를 얻을 수 있다.

워터마크의 강인성을 평가하기 위해 신호 변형 공격을 수행했다. 공격이 수행된 이미지에 3레벨 IDWT을 통해 주파수 도메인으로 변환하고 워터마크를 삽입했던 LL3t, LH3t, HL3t에 SVD 특이값 분해를 수행한다. 이를 통해 얻은 Sw를 이용하여 St를 재구성하고, 공격 강도 설정 시 이용했던 a 값을 이용하여 워터마크를 추출한다. LL3, LH3, HL3에서 각각 추출된 3개의 워터마크를 이용하여 최종 워터마크 이미지를 복원한다.

워터마크 복원 시에는 LH3, HL3 워터마크 데이터에 대해 중간값 계산을 수행하여 중간 워터마크 데이터로 통합한다. 이렇게 얻어진 중간 워터마크 데이터와 LL3 대역에서 추출된 워터마크를 가중 결합을 통해 결합한다. 중간 워터마크 데이터는 가중치를 0.3으로, LL3 대역의 워터마크 데이터는

가중치를 0.7로 설정했다.

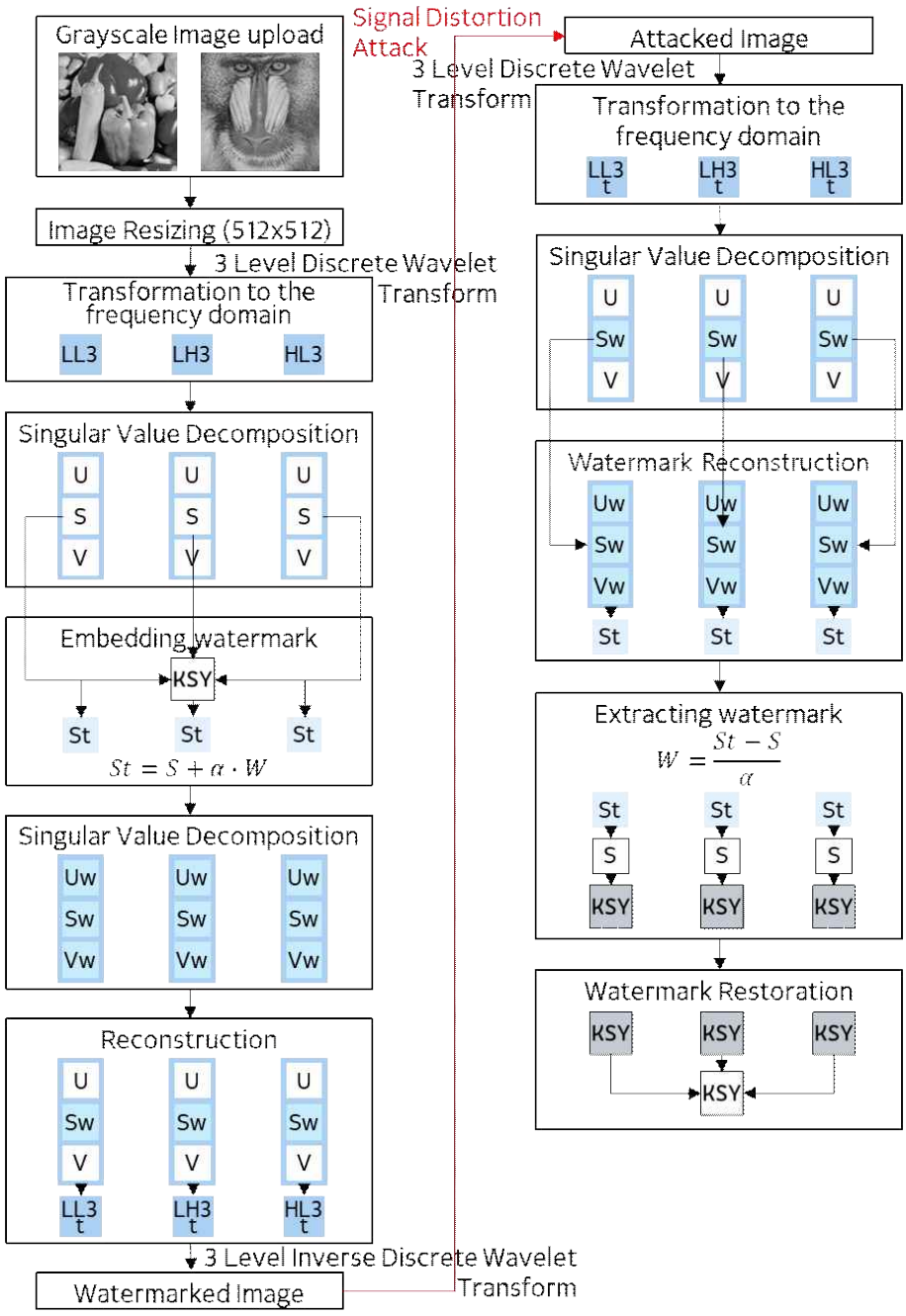


Figure 7. Flowchart of Experimental Process

5.3. 성능 평가 지표

워터마크 추출 성능 및 이미지 품질 비교를 위해 이용된 성능 평가 지표는 아래와 같다.

Normalized Cross-Correlation (NCC)는 두 이미지 간의 유사도를 측정하는 지표로 두 이미지가 얼마나 비슷한지를 나타낸다. 호스트 이미지와 워터마크가 삽입된 이미지 혹은 원본 워터마크와 추출된 워터마크 간의 유사성 평가를 위해 사용했다.

$$NCC = \frac{\sum^{i,j}(W(i,j) \cdot W'(i,j))}{\sqrt{\sum^{i,j} W(i,j)^2} \cdot \sqrt{\sum^{i,j} W'(i,j)^2}} \tag{2}$$

NCC는 수식 (2)과 같이 정의된다. $W(i,j)$는 호스트 이미지 또는 원본 워터마크를 의미하고, $W'(i,j)$는 워터마크가 삽입된 이미지 또는 추출된 워터마크를 의미한다. 두 이미지의 픽셀값 간 상관도와 이미지 각각의 에너지를 계산하여 정규화한 값을 나누어 계산한다. NCC의 결과는 -1에서 1 사이의 값을 갖게 되며, 1에 가까울수록 두 이미지가 유사하고 판단할 수 있다.

평균 제곱 오차(Mean Squared Error, MSE)는 원본 이미지와 변형된 이미지 간의 픽셀값 차이를 제곱한 후 평균을 구한 값으로 두 이미지 간의 수치적 오차를 측정한다. 워터마크 임베딩 전, 후의 픽셀값 차이를 평가할 수 있다.

$$MSE = \frac{1}{m \cdot n} \sum_{i=1}^{m} \sum_{j=1}^{n} [I(i,j) - K(i,j)]^2 \tag{3}$$

MSE는 수식 (3)와 같이 정의된다. $I(i,j)$는 원본 이미지, $K(i,j)$는 변형된 이미지의 픽셀값에 해당되고, m, n은 각 이미지의 크기이다. 값이 0에 가까울수록 두 이미지가 유사함을 의미한다.

신호 대 잡음비(Peak Signal-to-Noise Ratio, PSNR)는 원본 이미지와 변형이 수행된 이미지 간의 차이를 측정하는 지표로, 원본 이미지 대비 이미지 품질이 얼마나 유지되었는지 평가한다. 호스트 이미지에 워터마크를 임베딩한 후, 이미지 품질이 얼마나 잘 유지되었는지 평가하기 위해 이용된다.

$$PSNR = 10 \cdot log_{10}(\frac{MAX^2}{MSE}) \tag{4}$$

PSNR은 수식 (4)과 같이 정의된다. MAX는 픽셀값의 최대값이며, 보통 255로 나타난다. PSNR 값이 클수록 이미지 품질이 좋게 유지되었음을 의미하며, 일반적으로 30dB 이상일 경우 인간의 시각 체계로 이미지 품질 손상을 거의 인식하지 못한다.

구조적 유사도 지수(Structural Similarity Index Measure, SSIM)는 두 이미지 간의 구조적 유사성을 측정하는 지표로, 인간의 시각적 특성을 반영하여 이미지 품질을 평가한다. PSNR의 경우 수치를 기반으로 이미지 품질을 평가한다면, SSIM은 이미지의 구조적 특성과 밝기, 대비 등 인간이 인식하기 쉬운 차이를 고려하여 품질을 평가한다.

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{5}$$

SSIM은 수식 (5)과 같이 정의된다. 두 이미지의 평균(μ_x, μ_y), 분산(σ_x, σ_y), 공분산(σ_{xy}) 값을 이용하며, 안정성을 위해 작은 상수(C_1, C_2)를 포함한다. SSIM의 결과는 0과 1사이의 값을 가지게 되며, 0.7 이상인 경우 어느 정도 좋은 품질을 유지했다고 판단한다.

Ⅵ. 성능 평가

6.1 워터마크 반복 삽입에 따른 워터마킹 성능 비교

제안한 워터마킹 방식은 LL3, LH3, HL3 성분에 워터마크를 반복 삽입하여 총 3개의 워터마크가 삽입된다. 워터마크를 반복 삽입함에 따라 호스트 이미지의 품질에 미치는 영향과 워터마킹 추출 성능을 비교한다. 이를 위해 LL3 성분에 워터마크를 1회 삽입한 워터마킹 이미지와 제안 방식을 적용하여 LL3, LH3, HL3 성분에 각각 워터마크를 삽입하여 총 3회 삽입한 워터마킹 이미지를 비교한다. 워터마크 삽입 강도는 $a = 0.1$로 설정했다.

1) 이미지 품질 비교

Table 2와 Table 3은 각각 peppers와 mandrill 이미지에 대해 워터마크를 반복 삽입한 결과를 비교한 표이다. 두 이미지 모두 워터마크를 1회 삽입했을 때, 이미지 품질이 더 좋게 나타났다. LL3는 저주파 영역으로 이미지의 주요 정보가 포함되어 있으며, 변형이 크지 않으면 호스트 이미지의 시각적 품질에 큰 영향을 미치지 않는다. 워터마크를 3회 반복해서 삽입한 경우, 저주파 영역뿐만 아니라 고주파 영역(LH3, HL3)에도 삽입된 워터마크가 영향을 미쳐 약간의 이미지 품질 저하가 발생했다. 또한 동일한 워터마크를 반복적으로 임베딩하면서 변경된 계수가 누적되어 왜곡이 발생했을 가능성이 있다.

하지만 워터마크를 3회 반복하여 삽입했을 때 두 이미지 모두 PSNR 측면에서 42dB 이상의 결과를 보였는데, 이는 호스트 이미지와 워터마킹된

이미지가 거의 동일한 수준의 시각적 품질을 유지하고 있음을 의미한다. 또한 SSIM 측면에서도 98 이상의 결과를 보이며 단순한 픽셀 차이뿐 아니라 이미지의 전반적인 구조적 유사성 측면에서도 잘 유지되고 있음을 보여준다.

Table 2. Comparison of Image Quality Based on Watermark Repetition - Peppers

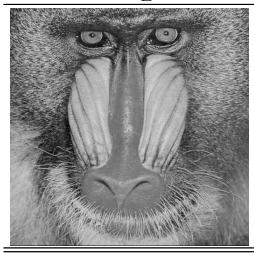
Watermarked Image	PSNR (unit:dB)	SSIM
	55.1247	0.9997

(a) 워터마크 1회 삽입

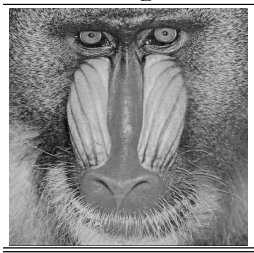
Watermarked Image	PSNR (unit:dB)	SSIM
	42.0135	0.9860

(b) 워터마크 3회 삽입

Table 3. Comparison of Image Quality Based on Watermark Repetition - Mandrill

Watermarked Image	PSNR (unit:dB)	SSIM
	53.7095	0.9999

(a) 워터마크 1회 삽입

Watermarked Image	PSNR (unit:dB)	SSIM
	42.0468	0.9955

(b) 워터마크 3회 삽입

2) 워터마크 추출 성능 비교

워터마크 반복 삽입에 따른 추출 성능을 비교하기 위해 워터마킹된 이미지에 신호 변형 공격을 수행한 후, 워터마크를 추출했다. 실험에는 7가지의 신호 변형 공격 기법이 고려되었다. 두 이미지 모두 워터마크를 3번 반복하여 삽입했을 때 NCC와 PSNR 측면에서 뛰어난 추출 성능을 보였다.

NCC는 워터마크의 추출 성능을 측정하는 지표로, 워터마크의 정확성과 회복력을 평가할 수 있다. 동일한 워터마크를 여러 영역(LL3, LH3, HL3)에 반복 삽입하면 한 영역에서 워터마크가 손상되더라도 다른 영역에서 워터마크 정보를 복구할 수 있으므로 추출된 워터마크와 원본 워터마크 간의 상관성을 개선할 수 있다.

PSNR은 호스트 이미지와 워터마킹된 이미지 간의 픽셀 차이를 기반으로 측정되기 때문에 3회 반복하여 삽입할 경우, 삽입 강도가 분산되어 변형 정도가 상대적으로 적어질 수 있고 픽셀값 변화의 극단적인 왜곡을 감소시킬 수 있다.

SSIM은 구조적 유사성을 평가하는 지표로 밝기, 대비, 구조 등의 요소를 기반으로 측정된다. 워터마크를 3회 반복하여 삽입하며 고주파 대역에도 워터마크가 삽입되는데, 삽입 위치가 다양해질수록 전체 구조적 유사성이 미치는 영향이 증가하게 된다. 반복 삽입을 통해 이미지의 텍스처나 패턴이 미세하게 변경될 수 있고, 이 과정에서 이미지의 구조적 정보가 손상될 수 있다. 이러한 이유로 워터마크를 3회 반복하여 삽입했을 때 워터마크를 1회 삽입한 경우보다 열화된 성능을 보였다.

Table 4. Comparison of Extraction Performance Based on Watermark Repetition - Peppers

		(a) 1회 반복		(b) 3회 반복	
Gaussian Noise	NCC		0.9729		0.9966
	PSNR		13.2802		29.5758
	SSIM		0.9265		0.8470
Salt & Pepper	NCC		0.7581		0.9948
	PSNR		2.7386		29.1978
	SSIM		0.9400		0.7985
Sparkle Noise	NCC		0.3807		0.9945
	PSNR		-6.6502		28.8502
	SSIM		0.9881		0.7690
JPEG Compression	NCC		0.9997		0.9998
	PSNR		32.8903		39.3492
	SSIM		0.9460		0.9574
JPEG2000 Compression	NCC		0.9958		0.9995
	PSNR		21.3219		36.7879
	SSIM		0.9637		0.9420
Blurring Attack	NCC		0.9278		0.9934
	PSNR		8.9221		28.1409
	SSIM		0.8452		0.8406
Low-frequency Filtering	NCC		0.7553		0.9907
	PSNR		2.7589		27.8305
	SSIM		0.8098		0.7685

Table 5. Comparison of Extraction Performance Based on Watermark Repetition - Mandrill

		(a) 1회 반복		(b) 3회 반복	
Gaussian Noise	NCC		0.8344		0.9744
	PSNR		14.2402		31.8009
	SSIM		0.9339		0.8666
Salt & Pepper	NCC		0.5815		0.9638
	PSNR		7.9314		28.6349
	SSIM		0.9615		0.7932
Sparkle Noise	NCC		0.1456		0.9715
	PSNR		-6.0371		28.9332
	SSIM		0.9961		0.8201
JPEG Compression	NCC		0.9971		0.9980
	PSNR		32.9038		39.9459
	SSIM		0.9353		0.9734
JPEG2000 Compression	NCC		0.7271		0.9903
	PSNR		10.9613		32.2040
	SSIM		0.9747		0.8641
Blurring Attack	NCC		0.6883		0.9327
	PSNR		10.0996		27.7186
	SSIM		0.8385		0.7924
Low-frequency Filtering	NCC		0.4268		0.8975
	PSNR		4.0761		27.5768
	SSIM		0.8015		0.6841

peppers 이미지의 경우, 워터마크를 1회 삽입했을 때, Sparkle Noise와

Low-frequency Filtering 공격에 의해 워터마크가 크게 훼손되었다. 워터마크를 3회 삽입한 경우, NCC 측면에서 평균 0.996, PSNR 측면에서 평균 31.390dB의 결과를 보이며 워터마크를 성공적으로 복원했다.

mandrill 이미지의 경우, Salt & Pepper, Sparkle Noise, Blurring, Low-frequency Filtering 공격에 의해 워터마크가 크게 훼손되었다. 워터마크를 3회 삽입했을 때 NCC 측면에서 0.961, PSNR 측면에서 평균 30.973dB의 결과를 보이며 워터마크가 잘 추출된 것을 확인할 수 있다.

6.2 종래 방식과 제안 방식의 워터마킹 성능 비교

제안 방식의 성능을 입증하기 위해 제안 방식과 종래 방식의 성능 비교를 수행했다. 종래 방식은 2레벨 DWT와 SVD를 결합한 이미지 워터마킹 방식을 제안했으며, 저주파 영역(LL2)에 워터마크를 임베딩했다[33]. 워터마크 삽입 강도는 종래 방식과 제안 방식 모두 $a = 0.1$로 설정했으며, 공격 강도를 5단계로 나누어 여러 강도의 공격에 대한 강인성을 평가했다. 공격 강도에 대한 세부 설정은 Table 1과 같다.

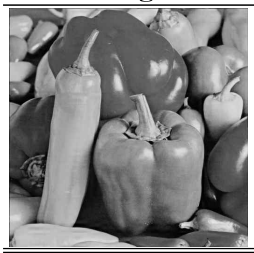
1) 이미지 품질 비교

Table 6, 7은 peppers와 mandrill 이미지에 대해 종래 방식과 제안 방식을 이용하여 워터마킹한 경우, 워터마킹된 이미지의 품질을 비교한 표이다. 두 이미지 모두 종래 방식을 적용했을 때 이미지 품질이 미세하게 더 우수하게 유지되는 경향을 확인할 수 있었다. 종래 방식의 경우, 저주파 영역(LL2)에만 워터마크를 임베딩하므로 이미지 품질이 더 우수하게 유지된다. PSNR 측면에서 peppers는 11.5%, mandrill은 7.28% 열화된 성능을 보였다. 하지만 여전히 PSNR 값은 40dB 이상으로 유지되었으며, 시각적 품질 평가 지표인 SSIM에서도 1.25%와 0.4%의 미미한 열화만 나타나면서 0.98 이상의 높은 SSIM 값을 기록하였다.

Table 6. Comparison of Image Quality Between Conventional and Proposed Methods - Peppers

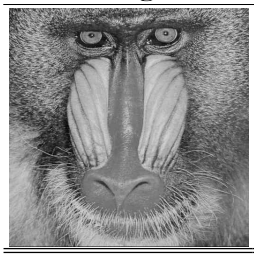
Watermarked Image	PSNR (unit:dB)	SSIM
	47.4871	0.9985

(a) 종래 방식

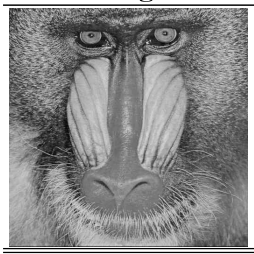
Watermarked Image	PSNR (unit:dB)	SSIM
	42.0135	0.9860

(b) 제안 방식

Table 7. Comparison of Image Quality Between Conventional and Proposed Methods - Mandrill

Watermarked Image	PSNR (unit:dB)	SSIM
	45.3471	0.9995

(a) 종래 방식

Watermarked Image	PSNR (unit:dB)	SSIM
	42.0468	0.9955

(b) 제안 방식

2) 워터마크 추출 성능 비교

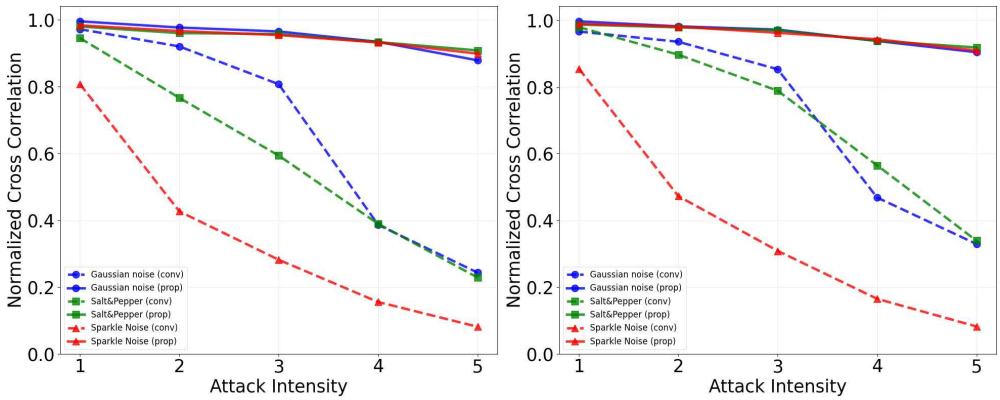
종래 방식과 제안 방식의 워터마크 추출 성능을 비교하기 위해 워터마킹 된 이미지에 신호 변형 공격을 수행한 후, 워터마크를 추출했다. 실험에는 7가지의 신호 변형 공격 기법이 고려되었다. 공격은 강도가 약한 공격부터

매우 강한 수준의 공격까지 5단계로 수행되었으며, 공격 강도에 따른 워터마크 추출 성능을 평가했다.

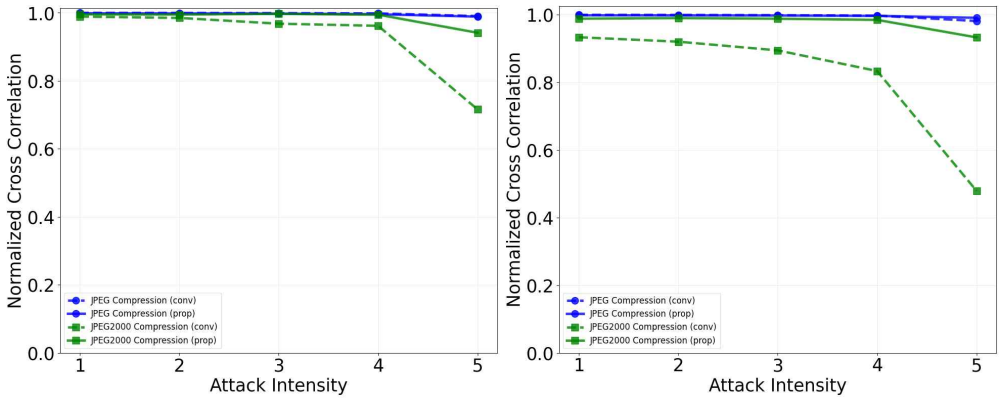
Figure 8은 Peppers 및 Mandrill 이미지를 대상으로 신호 변형 공격 강도에 따라 종래 방식과 제안 방식의 워터마킹 추출 성능을 NCC 관점에서 비교한 그래프이다.

종래 방식을 적용한 경우, 두 이미지 모두 가우시안 노이즈, 스파클 노이즈, 저주파 필터링 공격에서 공격 강도가 증가함에 따라 급격한 성능 열화를 보였으며, Salt & Pepper 및 블러링 공격에서도 뚜렷한 성능 저하가 관찰되었다. Peppers 이미지의 경우, 가우시안 노이즈, 스파클 노이즈, 저주파 필터링 공격을 수행했을 때 1단계에서 5단계로 공격 강도가 증가함에 따라 각각 75%, 89.99%, 82.55%의 성능 열화가 발생하였다. Mandrill 이미지에서는 각각 65.95%, 90.44%, 92.34%의 성능 열화가 나타났다.

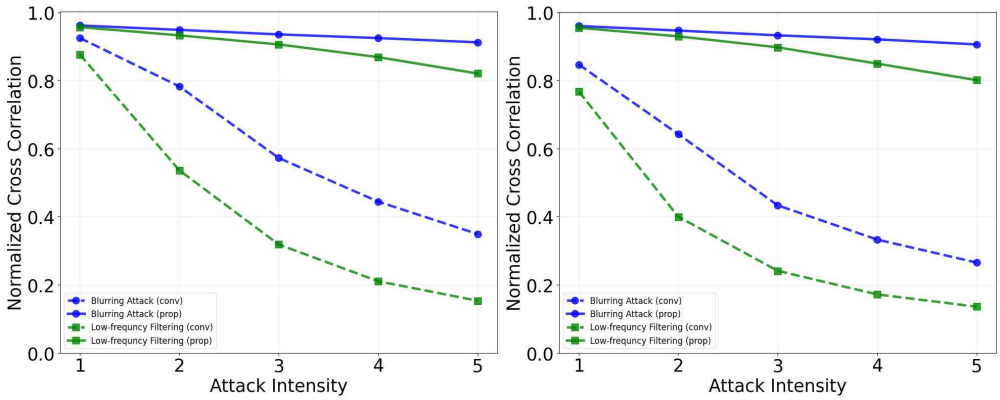
반면, 제안 방식을 적용했을 때 공격 강도 증가에 따른 성능 열화는 관찰되었으나, 그 정도가 미미한 수준에 그쳤다. 가장 큰 성능 열화가 발생한 저주파 필터링 공격에서도 Peppers 이미지가 14.26%, Mandrill 이미지가 16.10%의 성능 열화를 보이며, 이는 종래 방식과 비교했을 때 상대적으로 낮은 수준의 열화이다. 압축 공격에 대해서는 종래 방식과 제안 방식 간의 성능 차이가 크게 나타나지 않았으나, JPEG2000 공격을 5단계로 수행했을 때 제안 방식이 종래 방식 대비 Peppers 이미지에서 31.47%, Mandrill 이미지에서 94.66% 향상된 추출 성능을 기록하였다.



(a) Noise attack (from left: peppers, mandrill)



(b) Compression attack (from left: peppers, mandrill)



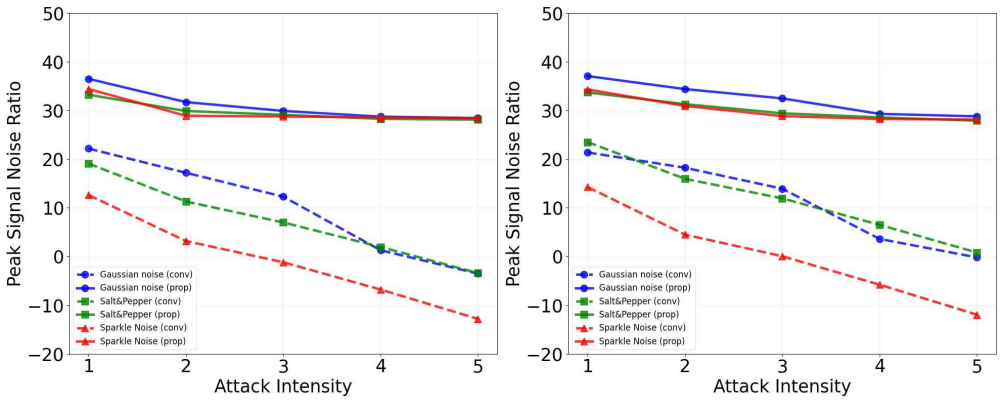
(c) Filtering attack (from left: peppers, mandrill)

Figure 8. Comparison of Extraction Performance of Conventional and Proposed Methods Based on Attack Intensity (NCC)

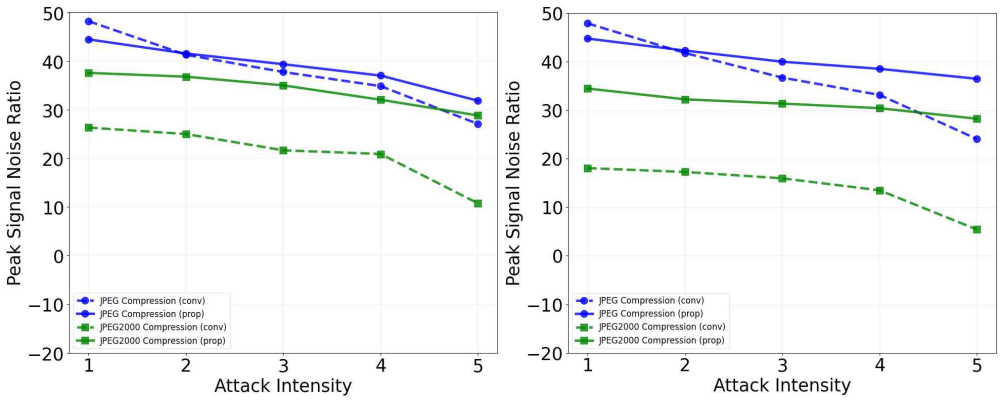
Figure 9는 Peppers 및 Mandrill 이미지를 대상으로 신호 변형 공격 강도에 따라 종래 방식과 제안 방식의 워터마킹 추출 성능을 PSNR 측면에서 비교한 그래프이다.

종래 방식을 적용한 경우, 두 이미지 모두 모든 공격에서 공격 강도가 증가함에 따라 워터마크 추출 성능이 크게 열화되는 경향을 보였다. 특히, JPEG 압축 공격을 제외하면 약한 수준의 공격인 1단계 공격에서도 추출 성능이 안정적으로 유지되지 못하는 것을 확인할 수 있었다. 성능 열화가 가장 두드러진 경우는 스파클 노이즈 공격으로, Peppers 이미지는 1단계 공격 수행 후 12.60dB에서 5단계 공격 수행 후 -12.83dB로 201.86%의 성능 열화가 발생했다. Mandrill 이미지도 1단계 공격 수행 후 14.25dB에서 5단계 공격 수행 후 -11.98dB로 184.05%의 심각한 성능 열화가 관찰되었다.

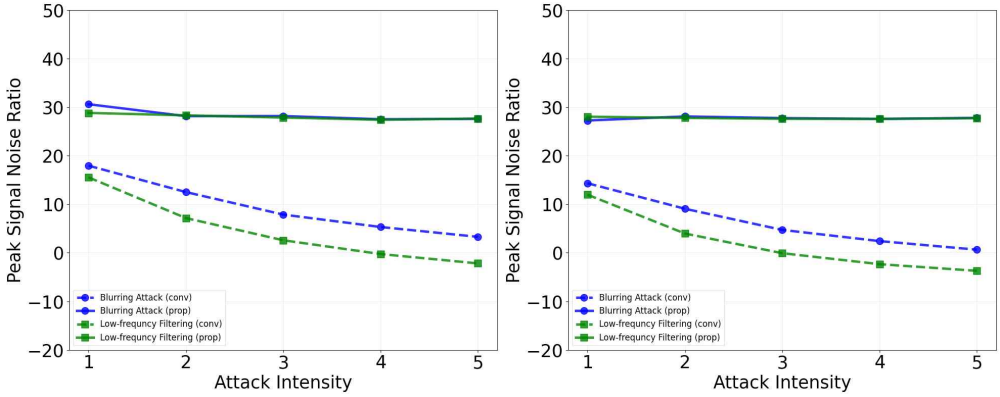
반면, 제안 방식을 적용하면 전체적으로 추출 성능이 안정적으로 유지되었다. Peppers 이미지에서는 가우시안 노이즈, Salt & Pepper, 스파클 노이즈, 저주파 필터링 공격 수행 시 공격 강도가 1단계에서 2단계로 증가함에 따라 약간의 성능 열화가 발생했으나 이후 단계에서는 유사한 성능을 유지하였다. 압축 공격의 경우에도 공격 강도 증가에 따른 성능 열화가 관찰되었으나, 종래 방식에 비해 열화 수준이 현저히 낮았으며, 5단계 공격 수행 후에도 더 높은 추출 성능을 나타냈다.



(a) Noise attack (from left: peppers, mandrill)



(b) Compression attack (from left: peppers, mandrill)



(c) Filtering attack (from left: peppers, mandrill)

Figure 9. Comparison of Extraction Performance of Conventional and Proposed Methods Based on Attack Intensity (PSNR)

3) 공격 강도에 따른 워터마킹된 이미지 품질 비교

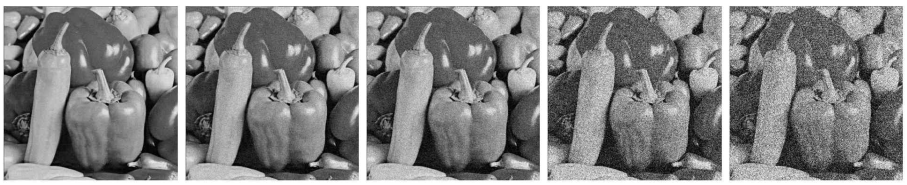
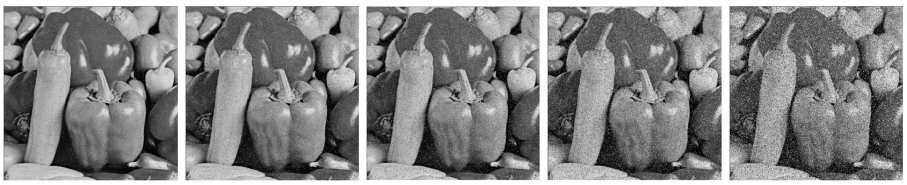
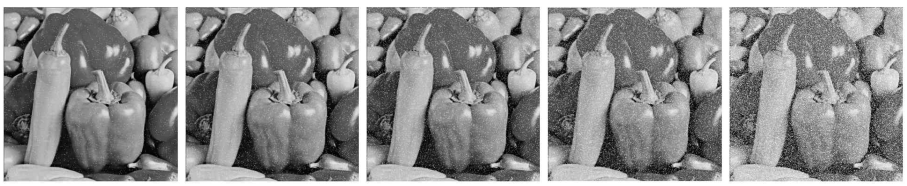
신호 변형 공격의 주요 목표는 워터마크가 포함된 이미지를 불법적으로 활용하기 위해 워터마크를 손상시켜 원본 이미지만을 남기는 데 있다. 그러나 원본 이미지에 과도한 손상이 발생할 경우, 워터마크를 효과적으로 제거하더라도 해당 이미지는 활용이 불가능하다.

Table 8, 9는 1레벨부터 5레벨의 공격이 수행되었을 때, 워터마킹된 이미지에 미치는 영향을 비교한 표이다. 각각 Peppers, Mandrill 이미지에 워터마크를 임베딩 한 후, 7가지 신호 변형 공격을 1레벨부터 5레벨까지 수행했으며, 워터마킹된 이미지와 공격이 수행된 이미지의 PSNR 및 SSIM 값을 비교했다.

pepeprs 이미지는 노이즈 공격이 수행되었을 때, PSNR 측면에서는 비교적 준수한 품질을 유지했지만 SSIM 측면에서 워터마킹된 이미지가 가장 많이 훼손되었다. 2단계 공격 수행 시 SSIM이 평균 0.46으로 나타났고 이후 3, 4, 5단계 공격에서도 평균 0.32, 017, 0.10으로 이미지 품질이 크게 훼손되었다.

mandrill 이미지는 필터링 공격이 수행되었을 때, 워터마킹된 이미지가 가장 많이 훼손되었으며, 가장 약한 공격이 수행되었을 경우(1레벨)에도 평균 PSNR이 30.31, 평균 SSIM이 0.77로 나타났다. 노이즈 공격의 경우, 다른 공격에 비해 PSNR 측면에서 준수한 품질을 유지했지만, SSIM 측면에서 열화가 크게 나타났다. 특히 3단계 공격이 수행되었을 때 SSIM 측면에서 평균 0.56으로 나타났으며, 3단계 이상의 공격이 수행된 경우 급격한 품질 열화가 발생했다.

Table 8. Comparison of Watermarked Image Quality Based on Attack Intensity - Peppers

	1 Level	2 Level	3 Level	4 Level	5 Level
Gaussian Noise					
	PSNR: 31.19 SSIM: 0.7	PSNR: 28.9 SSIM: 0.4	PSNR: 28.4 SSIM: 0.28	PSNR: 27.92 SSIM: 0.12	PSNR: 28.86 SSIM: 0.08
Salt & Pepper					
	PSNR: 47.78 SSIM: 0.76	PSNR: 43.0 SSIM: 0.49	PSNR: 40.81 SSIM: 0.34	PSNR: 37.94 SSIM: 0.19	PSNR: 35.14 SSIM: 0.1
Sparkle Noise					
	PSNR: 47.92 SSIM: 0.75	PSNR: 43.04 SSIM: 0.48	PSNR: 40.68 SSIM: 0.33	PSNR: 37.71 SSIM: 0.2	PSNR: 34.7 SSIM: 0.12

(a) Noise attack

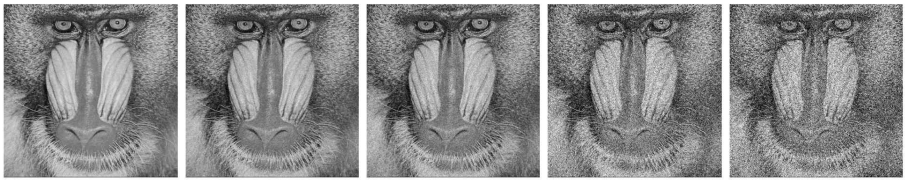
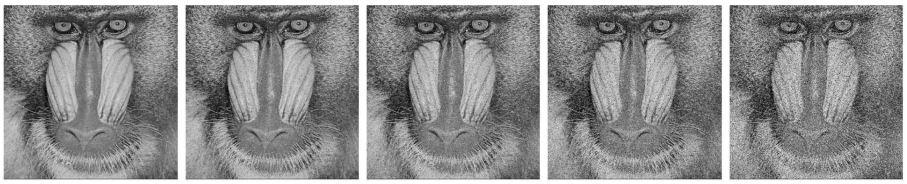
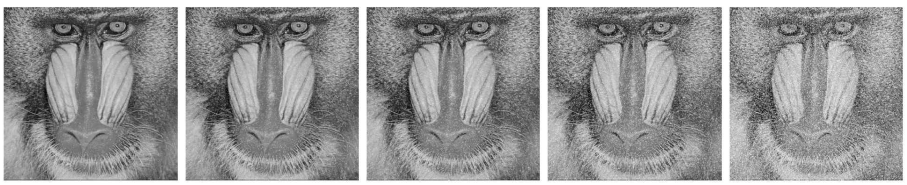
	1 Level	2 Level	3 Level	4 Level	5 Level
JPEG Compression					
	PSNR: 38.84 SSIM: 0.94	PSNR: 36.13 SSIM: 0.9	PSNR: 35.21 SSIM: 0.88	PSNR: 34.3 SSIM: 0.85	PSNR: 32.11 SSIM: 0.77
JPEG2000 Compression					
	PSNR: 36.23 SSIM: 0.9	PSNR: 35.63 SSIM: 0.88	PSNR: 34.86 SSIM: 0.86	PSNR: 33.76 SSIM: 0.84	PSNR: 31.38 SSIM: 0.74

(b) Compression attack

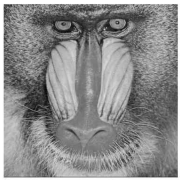
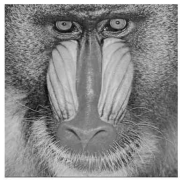
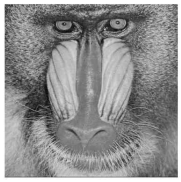
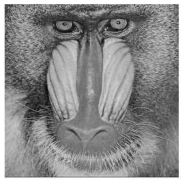
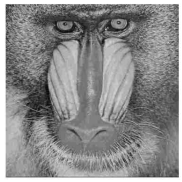
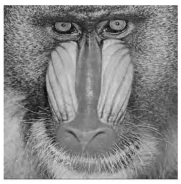
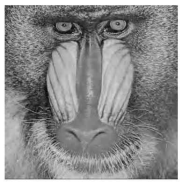
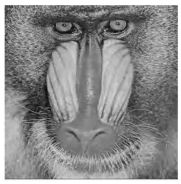
	1 Level	2 Level	3 Level	4 Level	5 Level
Blurring Attack					
	PSNR: 36.05 SSIM: 0.91	PSNR: 34.86 SSIM: 0.88	PSNR: 33.9 SSIM: 0.85	PSNR: 33.36 SSIM: 0.82	PSNR: 32.91 SSIM: 0.8
Low-Frequency Filtering Attack					
	PSNR: 35.13 SSIM: 0.88	PSNR: 33.57 SSIM: 0.82	PSNR: 32.67 SSIM: 0.77	PSNR: 32.01 SSIM: 0.72	PSNR: 31.51 SSIM: 0.68

(c) Filtering attack

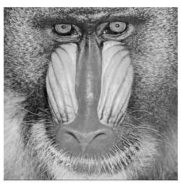
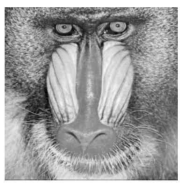
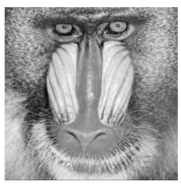
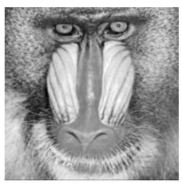
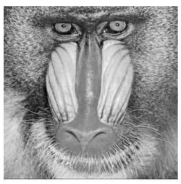
Table 9. Comparison of Watermarked Image Quality Based on Attack Intensity - Mandrill

	1 Level	2 Level	3 Level	4 Level	5 Level
Gaussian Noise					
	PSNR: 31.16 SSIM: 0.89	PSNR: 28.88 SSIM: 0.68	PSNR: 28.38 SSIM: 0.56	PSNR: 27.9 SSIM: 0.27	PSNR: 27.87 SSIM: 0.18
Salt & Pepper					
	PSNR: 48.02 SSIM: 0.86	PSNR: 43.25 SSIM: 0.68	PSNR: 41.07 SSIM: 0.55	PSNR: 38.21 SSIM: 0.39	PSNR: 35.42 SSIM: 0.24
Sparkle Noise					
	PSNR: 48.08 SSIM: 0.87	PSNR: 43.19 SSIM: 0.68	PSNR: 41.0 SSIM: 38.03	PSNR: 38.03 SSIM: 0.41	PSNR: 34.96 SSIM: 0.26

(a) Noise attack

	1 Level	2 Level	3 Level	4 Level	5 Level
JPEG Compression					
	PSNR: 37.1 SSIM: 0.98	PSNR: 32.0 SSIM: 0.93	PSNR: 31.05 SSIM: 0.9	PSNR: 30.43 SSIM: 0.85	PSNR: 29.49 SSIM: 0.71
JPEG2000 Compression					
	PSNR: 30.21 SSIM: 0.82	PSNR: 29.86 SSIM: 0.77	PSNR: 29.57 SSIM: 0.71	PSNR: 29.27 SSIM: 0.62	PSNR: 28.94 SSIM: 0.43

(b) Compression attack

	1 Level	2 Level	3 Level	4 Level	5 Level
Blurring Attack					
	PSNR: 30.57 SSIM: 0.83	PSNR: 29.97 SSIM: 0.72	PSNR: 29.59 SSIM: 0.6	PSNR: 29.43 SSIM: 0.53	PSNR: 29.3 SSIM: 0.48
Low-Frequency Filtering Attack					
	PSNR: 30.04 SSIM: 0.72	PSNR: 29.42 SSIM: 0.51	PSNR: 29.18 SSIM: 0.4	PSNR: 29.03 SSIM: 0.34	PSNR: 28.92 SSIM: 0.3

(c) Filtering attack

6.3 워터마크 삽입 강도에 따른 워터마킹 성능 비교 분석

제안한 워터마킹 방식은 워터마크 임베딩 시 파라미터 a를 통해 워터마크 삽입 강도를 설정할 수 있다. a값이 클수록 워터마크가 강하게 삽입되며, 워터마크의 강인성을 향상시킬 수 있지만, 호스트 이미지의 품질이 훼손될 수 있다. 제안 방식에 대한 다양한 조건의 성능을 비교 분석하여 워터마크 삽입 강도에 따른 이미지 품질 및 워터마크 강인성에 대해 평가하기 위해 $a = [0.1, 0.15, 0.2, 0.25, 0.3]$으로 조정하며 실험을 수행했다. 실험을 peppers, mandrill 이미지에 대해 수행되었으며, 앞서 언급한 7개의 신호 변형 공격 기법을 이용하여 워터마크의 강인성을 평가했다. 공격 강도는 중간 강도의 공격인 3레벨로 설정했으며, 각 공격에 대한 파라미터 설정값은 Table 1에 언급되어 있다.

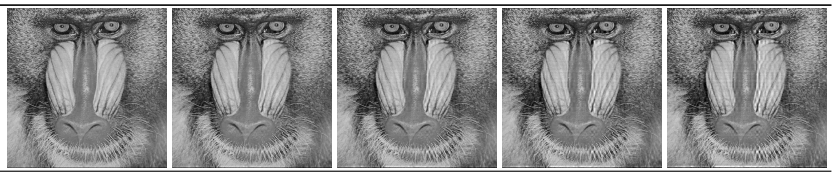
1) 이미지 품질 비교

Table 10, 11은 각각 peppers와 mandrill 이미지에 대해 워터마킹 강도를 조정하여 워터마크를 임베딩한 결과를 비교한 표이다. 두 이미지 모두 워터마킹 임베딩 강도가 높을수록 호스트 이미지의 품질이 저하되는 것을 볼 수 있다.

Table 10. Comparison of Image Quality Based on Watermark Embedding Strength - Peppers

Alpha(α)	0.1	0.15	0.2	0.25	0.3
Watermarked Image					
PSNR	42.0135	37.8787	35.7306	34.2976	33.2987
SSIM	0.9860	0.9631	0.9361	0.9074	0.8791
MSE	4.0900	10.5977	17.3788	24.1721	30.4234

Table 11. Comparison of Image Quality Based on Watermark Embedding Strength - Mandrill

Alpha(α)	0.1	0.15	0.2	0.25	0.3
Watermarked Image					
PSNR	42.0469	37.8363	35.5500	33.9997	32.8573
SSIM	0.9956	0.9862	0.9735	0.9589	0.9434
MSE	4.0587	10.7017	18.1166	25.8883	33.6776

PSNR은 이미지 간의 픽셀값 차이를 기반으로 평가하기 때문에 두 이미지의 결과가 유사하게 나타났다. 워터마킹 삽입 강도가 높아지면서 변형되는 주파수 값의 범위가 증가하여 PSNR 측면의 열화가 발생하였다.

SSIM의 경우 이미지의 구조적 요소의 변화에 민감하다. Peppers 이미지의 경우 부드러운 곡선과 단순한 색상 전환이 많은데, 이러한 특징은 저주파 성분의 미세한 변화에 쉽게 영향을 받을 수 있다. Mandrill 이미지의 경

우 복잡한 텍스처와 모서리, 세부 패턴 등 고주파 성분이 풍부하다. 워터마크가 강하게 임베딩되어 일부 구조에 변형이 발생하더라도 본래 텍스처가 복잡하기 때문에 구조적 유사성에 미치는 영향이 상대적으로 덜하다. 이와 같은 이유로 Peppers 이미지가 Mandrill 이미지와 비교했을 때, SSIM 측면에서 더 큰 열화를 보였다.

2) 워터마크 추출 성능 비교

워터마크의 임베딩 강도에 따른 추출 성능을 비교하기 위해 워터마킹된 이미지에 신호 변형 공격을 수행한 후, 워터마크를 추출했다. 실험에는 7가지의 신호 변형 공격 기법이 고려되었다.

Figure 10, 11은 peppers. mandrill 이미지에 대해 워터마크 임베딩 강도에 따른 워터마크 추출 성능을 비교한 그래프이다. 노이즈 공격과 필터링 공격을 수행했을 때, 워터마크 임베딩 강도를 높게 설정할수록 워터마크 추출 성능이 개선된 것을 확인할 수 있다.

JPEG 및 JPEG2000 압축은 저주파 대역의 데이터를 최대한 유지하는 방향으로 동작하지만, 계수의 상대적 크기가 커질 경우 압축 알고리즘의 양자화 및 비트 할당 과정에서 워터마크 정보가 잘리거나 왜곡될 가능성이 높아진다. JPEG 압축의 경우, 임베딩 강도가 높아질수록 워터마크 정보의 비율이 증가하는데, 이는 양자화에 의해 불균형하게 제거될 가능성이 높아진다. JPEG2000 압축의 경우, 임베딩 강도가 높아지면 DWT 계수의 비율이 변화하기 때문에, 압축 과정에서 비트 분배가 비효율적으로 이루어질 가능성을 높인다. 이로 인해 압축 과정에서 워터마킹 데이터가 더 많이 손실될 수 있다.

필터링 공격의 경우, 워터마킹 임베딩 강도에 따라 워터마크 추출 성능에 가장 큰 개선을 보였다. peppers 이미지의 경우, NCC 측면에서 블러링 공격이 수행되었을 때 2.19%, 저주파 필터링 공격이 수행되었을 때 3.09%의 개선을 보였다. mandrill 이미지의 경우, NCC 측면에서 블러링 공격이 수행되었을 때 2.40%, 저주파 필터링 공격이 수행되었을 때 3.69%의 개선을 보였다. 추가로 실험을 진행하여 필터링 공격을 다양한 강도로 수행하고 워터마킹 임베딩 강도에 따른 워터마킹 추출 성능을 비교했다.

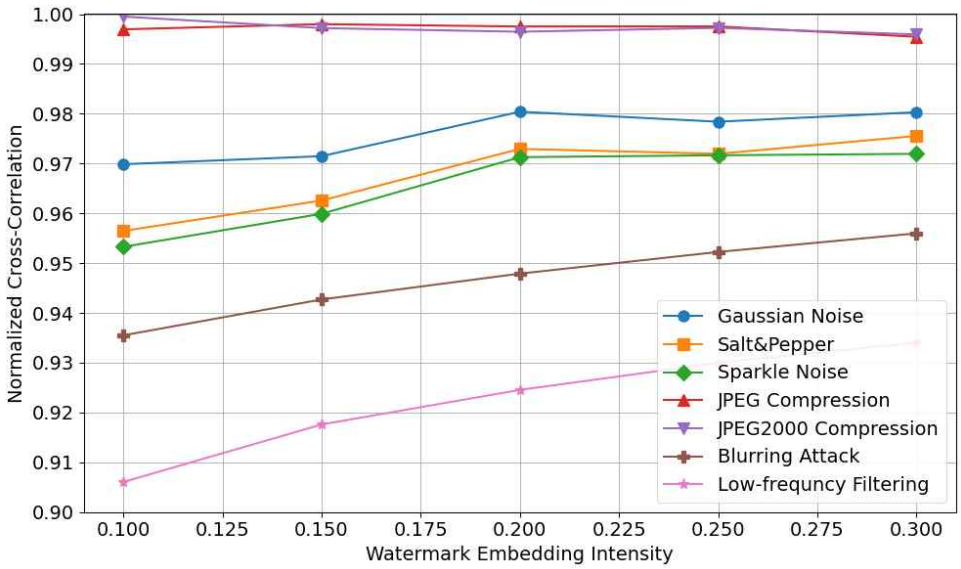


Figure 10. Comparison of Extraction Performance Based on Watermark Embedding Strength - Peppers

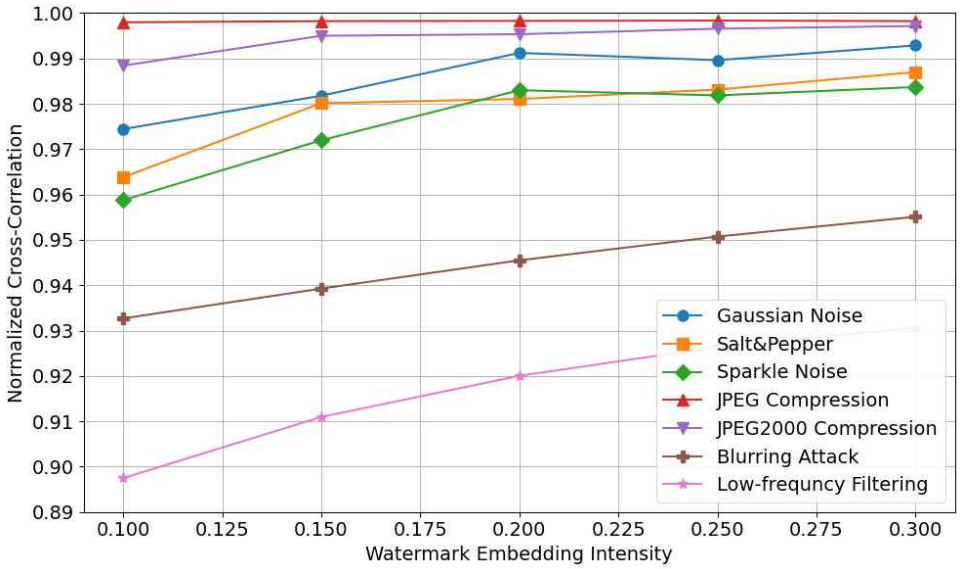
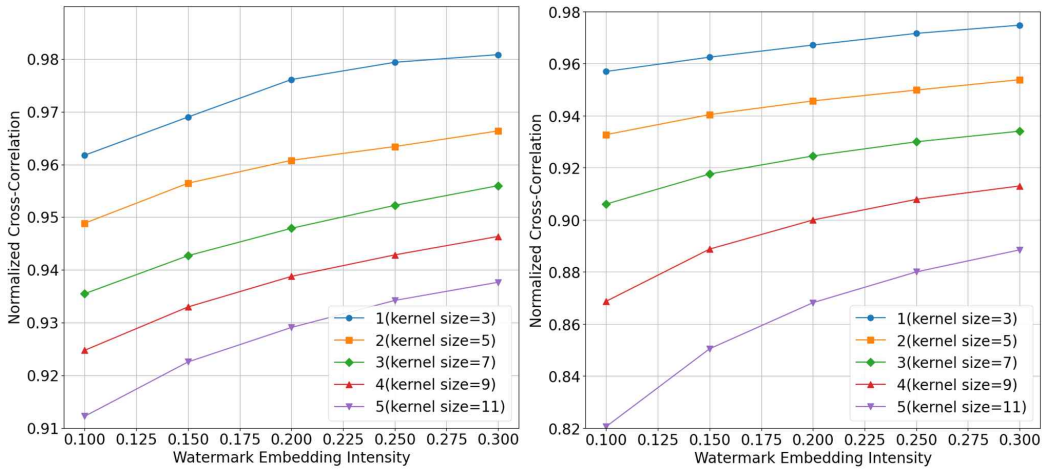


Figure 11. Comparison of Extraction Performance Based on Watermark Embedding Strength - Mandrill

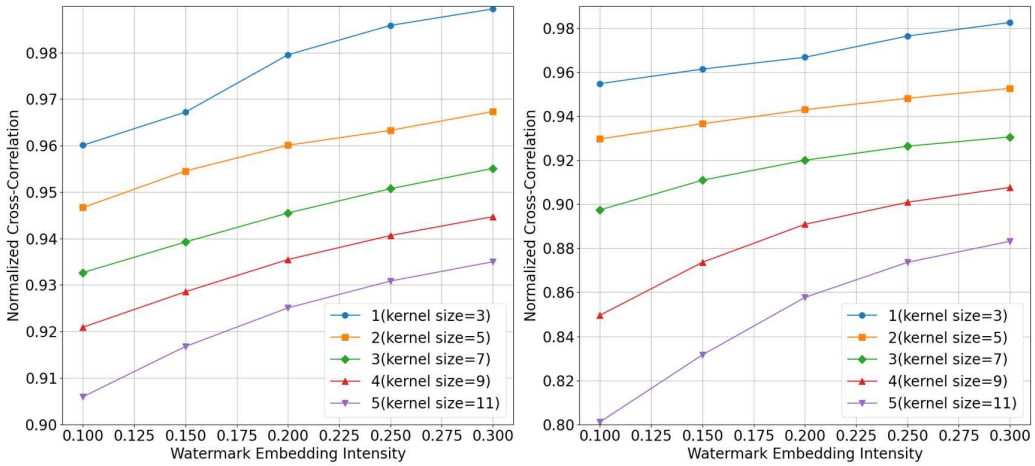
Figure 12, 13은 peppers와 mandrill 이미지에 필터링 공격을 다양한 강도로 수행하고 워터마크 임베딩 강도에 따라 추출 성능을 비교한 그래프이다. Table 1에 명시된 대로 공격 강도를 1에서 5로 조정했으며, 워터마킹 삽입 강도는 앞서 수행된 실험과 동일하게 $a = [0.1, 0.15, 0.2, 0.25, 0.3]$로 조정했다.

두 이미지 모두 공격 강도가 강해질수록 전체적인 추출 성능이 열화되었다. 동시에 워터마킹 삽입 강도가 강해질수록 추출 성능이 개선되는 것을 확인할 수 있다. 저주파 필터링 공격의 경우, 공격 강도를 5로 설정한 경우(Kernel size=11)에 전체적인 추출 성능이 열화 되었지만 워터마크 삽입 강도가 강해짐에 따라 가장 큰 성능 향상을 보였다. 공격 강도를 1로 설정한 경우(Kernel size=3)와 비교했을 때, peppers 이미지는 최대 8.27%, mandrill 이미지는 최대 10.25%의 추출 성능 개선이 이루어졌다.



(a) 블러링 공격 (b) 저주파 필터링 공격

Figure 12. Comparison of Extraction Performance Based on Attack Intensity and Watermark Embedding Strength - Peppers



(a) 블러링 공격 (b) 저주파 필터링 공격

Figure 13. Comparison of Extraction Performance Based on Attack Intensity and Watermark Embedding Strength - Mandrill

Ⅶ. 결론 및 향후 연구

본 논문에서는 신호 변형 공격에 강인하면서도 비가시성이 뛰어난 디지털 이미지 워터마킹 기법을 제안하였다. 제안한 기법은 3레벨 DWT와 SVD를 결합하여 저주파 영역(LL3)뿐만 아니라 일부 고주파 영역(LH3, HL3)에도 SVD를 수행한 후, 특이값에 워터마크를 삽입하고 이를 반복적으로 임베딩함으로써 이미지 품질의 저하를 최소화하면서도 다양한 신호 변형 공격에 대한 강인성을 확보할 수 있었다. 저주파 영역 외에도 고주파 영역에 워터마크를 반복 삽입하여, 손상된 신호를 상호 보완적으로 복원할 수 있는 구조를 구현하였다. 이러한 방식은 기존 연구에서 흔히 발생하는 비가시성과 강인성 간의 트레이드 오프(Trade-off) 문제를 효과적으로 완화하였다. 실험 결과에 따르면 제안된 워터마킹 방식은 호스트 이미지에 워터마크를 삽입한 후에도 높은 PSNR과 SSIM 값을 유지하며, 이미지 품질을 우수하게 보존하는 것으로 나타났다. 또한, 워터마크를 한 번만 삽입했을 때와 비교하여 반복 삽입을 통해 추출 성능이 크게 향상된 것을 확인할 수 있었다. 이를 통해 반복 삽입된 워터마크가 노이즈 및 신호 변형으로 손상되더라도 복원 과정에서 상호 보완적인 역할을 수행하여, 높은 신뢰도로 워터마크를 검출할 수 있음을 확인하였다. 또한 다양한 강도의 신호 변형 공격을 5단계로 조정하여 기존 방식과 제안 방식을 비교한 결과, 제안 방식은 노이즈 공격, 압축 공격 등 다양한 강도의 공격에도 워터마크를 성공적으로 추출하며 강인성을 입증하였다.

본 연구는 디지털 콘텐츠 보호에 있어 실질적인 활용 가능성을 제시하며, 높은 비가시성과 신뢰도가 요구되는 분야에서 효과적으로 적용될 수 있을 것으로 기대된다.

참 고 문 헌

[1] Zirpoli, C.T. Generative artificial intelligence and copyright law. 2023.

[2] Kahveci, Z.Ü. Attribution problem of generative AI: a view from US copyright law. Journal of Intellectual Property Law and Practice. 2023, 18(11), pp.796-807.

[3] Yao, H.; Lou, J.; Qin, Z.; Ren, K. Promptcare: Prompt copyright protection by watermark injection and verification. In Proceedings of the 2024 IEEE Symposium on Security and Privacy (SP), IEEE, May 2024, pp.845-861.

[4] Awasthi, D.; Tiwari, A.; Khare, P.; Srivastava, V.K. A comprehensive review on optimization-based image watermarking techniques for copyright protection. Expert Systems with Applications. 2024, 242, 122830.

[5] Kadian, P.; Arora, S.M.; Arora, N. Robust digital watermarking techniques for copyright protection of digital data: A survey. Wireless Personal Communications. 2021, 118, pp.3225-3249.

[6] Su, Q.; Zhang, X.; Wang, H. A blind color image watermarking algorithm combined spatial domain and SVD. International Journal of Intelligent Systems. 2022, 37(8), pp.4747-4771.

[7] Dharmika, B.; Kiran, V.; Muralidhar, A. Privacy protection of digital information using frequency domain watermarking technique. In Proceedings of the 2021 4th International Conference on Recent Trends in Computer Science and Technology (ICRTCST), IEEE, 2022,

pp.123-130.

[8] Zainol, Z.; Hani, M.A.; Hussain, A. Hybrid SVD-based image watermarking schemes: a review. IEEE Access. 2021, 9, pp.32931-32968.

[9] Zermi, N.; Ghozzi, A.; Hamza, I. Robust SVD-based schemes for medical image watermarking. Microprocessors and Microsystems. 2021, 84, 104134.

[10] Evsutin, O.; Dzhanashia, K. Watermarking schemes for digital images: Robustness overview. Signal Processing: Image Communication. 2022, 100, 116523.

[11] Zermi, N.; Ghozzi, A.; Hamza, I. A DWT-SVD based robust digital watermarking for medical image security. Forensic Science International. 2021, 320, 110691.

[12] Deepika, K.; Punj, D.; Jyoti. Spatial Domain Method for Image Analysis: A Grey-Level Computation Approach. In Proceedings of the International Conference on Advances in Computing and Data Sciences, Cham: Springer Nature Switzerland, 2023.

[13] Sheng, Z.; Lin, J.; Wang, H. Frequency-domain deep guided image denoising. IEEE Transactions on Multimedia. 2022, 25, pp.6767-6781.

[14] Naffouti, S.E.; Kricha, A.; Sakly, A. A sophisticated and provably grayscale image watermarking system using DWT-SVD domain. The Visual Computer. 2023, 39(9), pp.4227-4247.

[15] Xiong, L.; Zhong, X.; Yang, C.-N. DWT-SISA: A secure and effective discrete wavelet transform-based secret image sharing with authentication. Signal Processing. 2020, 173, 107571.

[16] Yasmeen, F.; Uddin, M.S. An efficient watermarking approach based on LL and HH edges of DWT - SVD. SN Computer Science. 2021, 2(2), 82.

[17] Analysis and effectiveness of deeper levels of SVD on performance of hybrid DWT and SVD watermarking. Multimedia Tools and Applications. 2024, 83(2), pp.3895-3916.

[18] Chen, R.; Zhou, M.; Li, F. Image denoising algorithm based on improved K singular value decomposition and atom optimization. CAAI Transactions on Intelligence Technology. 2022, 7(1), pp.117-127.

[19] Bose, A.; Maity, S.P. Secure sparse watermarking on DWT-SVD for digital images. Journal of Information Security and Applications. 2022, 68, 103255.

[20] Taj, R.; Khan, M.; Ashraf, R.; et al. A SURF and SVD-based robust zero-watermarking for medical image integrity. PloS One. 2024, 19(9), e0307619.

[21] Kahlessenane, F.; Kricha, A.; Sakly, A. A robust blind medical image watermarking approach for telemedicine applications. Cluster Computing. 2021, 24(3), pp.2069-2082.

[22] Wan, W.; Zhou, X.; Jiang, M. A comprehensive survey on robust image watermarking. Neurocomputing. 2022, 488, pp.226-247.

[23] Guan, Q.; Deng, H.; Liang, W.; Zhong, X.; Ma, M. Multi-images encryption and watermarking with small number of keys via computational ghost imaging. Optics & Laser Technology. 2024, 168, 109957.

[24] Zhao, Z.; Pan, M.; Xie, S. Model-driven deep unrolling: Towards

interpretable deep learning against noise attacks for intelligent fault diagnosis. ISA Transactions. 2022, 129, pp.644-662.

[25] Zhang, Q.; Zheng, S.; Wang, Z. Salt and pepper noise removal method based on graph signal reconstruction. Digital Signal Processing. 2023, 135, 103941.

[26] Limsuebchuea, A.; Duangsoithong, R.; Phukpattaranont, P. Self-Augmented Noisy Image for Noise2Noise Image Denoising. IEEE Access. 2024, 12, pp.71076-71087.

[27] Mehta, D.; Bhatti, D. Blind image steganography algorithm development which resistant against JPEG compression attack. Multimedia Tools and Applications. 2022, 81(1), pp.459-479.

[28] Natsheh, S.; Farajallah, M. Classification of jpeg2000 image encryption algorithms. J Data Analytic Eng Decision Making. 2024, 1(1), pp.01-09.

[29] Chen, H.; Zhu, T.; Zhao, Y.; Liu, B.; Yu, X.; Zhou, W. Low-frequency image deep steganography: Manipulate the frequency distribution to hide secrets with tenacious robustness. arXiv preprint arXiv:2303.13713. 2023.

[30] Sun, B.; Ma, X.; Wang, H. Defending Against Local Adversarial Attacks through Empirical Gradient Optimization. Tehnički vjesnik. 2023, 30(6), pp.1888-1898.

[31] Hosseini, S.A.; Farahmand, P. An attack resistant hybrid blind image watermarking scheme based on combination of DWT, DCT and PCA. Multimedia Tools and Applications. 2024, 83(7), pp.18829-18852.

[32] Lidyawati, L.; Kricha, A.; Sakly, A. Digital watermarking image

using three-level discrete wavelet transform under attacking noise. Bulletin of Electrical Engineering and Informatics. 2022, 11(1), pp.231-238.

[33] Kusumaningrum, D.P.; Asvini, T.; Prasetyo, W. DWT-SVD Combination Method for Copyrights Protection. Sci J Inform. 2020, 7(1), pp.311.

[34] Varun, K.S.; Mandava, A.K.; Chowdary, R. Robust DWT-SVD domain image watermarking based on iterative blending. Journal of Physics: Conference Series. 2021, Vol. 2070, No. 1, IOP Publishing.

ABSTRACT

Invisible Watermarking Using Discrete Wavelet Transform and Singular Value Decomposition

Seo-Yi Kim
Department of Future Convergence Technology Engineering
Graduate School of
Sungshin Women's University

With the advancement of AI technology, the issue of copyright infringement in digital images has become increasingly severe, drawing attention to digital watermarking as a key solution. Digital watermarking is a technical method that embeds and detects specialized watermarks to protect the copyright of digital content, or to detect and track forgery and tampering. A watermark must ensure imperceptibility, maintaining the visual quality of the original image, while also being robust against external factors such as image manipulation and noise attacks, allowing for stable extraction. However, previous studies have failed to resolve the trade-off issue where enhancing imperceptibility reduces robustness, or improving robustness degrades image quality.

This study proposes a method where a 3-level DWT is performed on an image to separate frequency components into multiple levels, followed by SVD applied across multiple regions to repeatedly embed a watermark into singular values. The objective is to achieve watermarking with high imperceptibility and robustness against signal manipulation attacks.

ACKNOWLEDGEMENTS

본 논문의 연구를 지도해 주신 이일구 교수님께 깊이 감사드리며, 소중한 조언과 격려를 아끼지 않으신 김성민 교수님과 임연섭 교수님께도 진심으로 감사드립니다.