(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0135302
(43) 공개일자 2024년09월10일

(51) 국제특허분류(Int. Cl.)
G06T 1/00 (2006.01) *G06N 3/0464* (2023.01)
G06N 3/09 (2023.01) *G06T 5/40* (2024.01)
G06V 10/40 (2022.01)
(52) CPC특허분류
G06T 1/0021 (2013.01)
G06N 3/0464 (2023.01)
(21) 출원번호 10-2023-0028710
(22) 출원일자 2023년03월03일
심사청구일자 2024년05월02일

(71) 출원인
국립한밭대학교 산학협력단
대전광역시 유성구 동서대로 125 (덕명동)
(72) 발명자
장한얼
대전광역시 유성구 동서대로 125 한밭대학교, N4동(산업정보관) 607호
김민준
대전광역시 유성구 지족동로 124 노은리슈빌3, 107-1402
(74) 대리인
특허법인팬브릿지

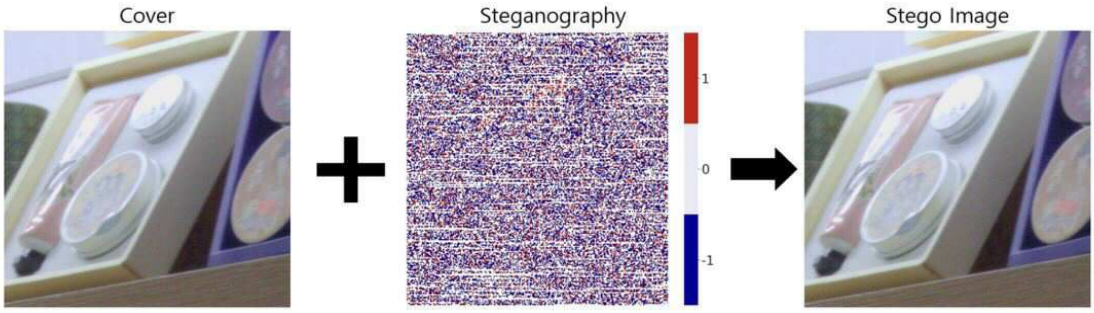
전체 청구항 수 : 총 6 항

(54) 발명의 명칭 **픽셀 히스토그램을 활용한 효율적인 스테가노그래피 탐지 방법**

(57) 요 약

픽셀 히스토그램을 이용한 스테가노그래피 탐지방법이 개시된다. 본 명세서는, 이미지에서 픽셀 히스토그램 벡터를 추출하고, 이를 사전에 학습된 다층퍼셉트론 기반의 네트워크를 이용하여 학습된 스테가노그래피 탐지모델에 입력하여, 커버 이미지와 스테고 이미지로 분류할 수 있다. 이에 따라 종래의 딥러닝 기반의 스테가노그래피 탐지 알고리즘에 비해 모델의 경량화, 추론 시간 감소 및 정확도면에서 우수한 성능을 구현할 수 있다.

대 표 도 - 도2



(52) CPC특허분류
G06N 3/09 (2023.01)
G06T 5/40 (2024.01)
G06V 10/40 (2023.08)
G06T 2201/0065 (2013.01)
G06T 2207/20081 (2013.01)
G06T 2207/20084 (2013.01)

이 발명을 지원한 국가연구개발사업
과제고유번호 1345356233
과제번호 2021RIS-004
부처명 교육부
과제관리(전문)기관명 한국연구재단
연구사업명 지자체대학협력기반지역혁신사업
연구과제명 (대전세종충남지역혁신플랫폼)충남대학교
기 여 율 1/1
과제수행기관명 (대전세종충남지역혁신플랫폼)충남대학교
연구기간 2022.05.01 ~ 2023.03.31

청구범위유예 : 있음
임시명세서출원 : 있음

명 세 서

청구범위

청구항 1

이미지를 수신하는 단계;

상기 이미지에서 픽셀 히스토그램 벡터를 추출하는 단계;

상기 픽셀 히스토그램 벡터를 상기 이미지의 특징벡터로 정의하여 사전에 학습된 스테가노그래피 탐지모델에 입력하는 단계; 및

상기 스테가노그래피 탐지모델의 출력에 기초하여 상기 이미지를 커버(cover) 이미지와 스테고(stego) 이미지로 분류하는 단계;를 포함하고,

상기 스테가노그래피 탐지모델은, 다층퍼셉트론(MLP, Multi-Layer Perceptron) 기반의 네트워크인 것을 특징으로 하는 픽셀 히스토그램을 이용한 스테가노그래피 탐지방법.

청구항 2

제 1 항에 있어서,

상기 스테가노그래피 탐지모델의 학습 데이터는, 상기 커버 이미지와 제1 레이블 데이터 쌍, 상기 스테고 이미지와 제2 레이블 데이터 쌍으로 구성되는 것을 특징으로 하는 픽셀 히스토그램을 이용한 스테가노그래피 탐지방법.

청구항 3

제 2 항에 있어서,

상기 스테고 이미지의 제2 픽셀 히스토그램에서 상기 커버 이미지의 제1 픽셀 히스토그램과 진동 폭에서 차이가 존재하는 범위의 데이터들이 상기 학습 데이터로 이용되는 것을 특징으로 하는 픽셀 히스토그램을 이용한 스테가노그래피 탐지방법.

청구항 4

제 1 항에 있어서,

상기 픽셀 히스토그램 벡터를 추출하는 단계는,

상기 픽셀 히스토그램의 값들 중 양 끝단의 값이 제거된 채 추출되는 값인 것을 특징으로 하는 픽셀 히스토그램을 이용한 스테가노그래피 탐지방법.

청구항 5

제 1 항에 있어서,

상기 픽셀 히스토그램 벡터는 상기 이미지의 RGB 채널 중 적어도 하나의 채널의 픽셀 히스토그램 벡터인 것을 특징으로 하는 픽셀 히스토그램을 이용한 스테가노그래피 탐지방법.

청구항 6

스테가노그래피 이미지를 탐지하는 장치로서,

적어도 하나의 프로그램이 기록된 메모리; 및

상기 프로그램을 실행하는 프로세서;를 포함하되,

상기 프로그램은,

수신된 이미지에서 픽셀 히스토그램 벡터를 추출하도록 하는 명령,

상기 픽셀 히스토그램 벡터를 상기 이미지의 특징벡터로 정의하여 사전에 학습된 스테가노그래피 탐지모델에 입력하도록 하는 명령을 포함하고,

상기 스테가노그래피 탐지모델은, 다층퍼셉트론(MLP, Multi-Layer Perceptron) 기반의 네트워크이며, 상기 스테가노그래피 탐지모델의 출력에 기초하여 상기 이미지는 커버(cover) 이미지와 스테고(stego) 이미지로 분류되는 것을 특징으로 하는 장치.

발명의 설명

기 술 분 야

[0001] 본 명세서는 스테가노그래피 탐지방법에 관한 것으로서, 보다 구체적으로 픽셀 히스토그램을 활용한 스테가노그래피 탐지방법에 관한 것이다.

배 경 기 술

[0002] 스테가노그래피는 기밀 메시지를 숨기는 기술로, 다른 사람들이 메시지의 존재를 인식하지 못하도록 하면서도 메시지를 전달하는 비밀 통신 기술이다. 컴퓨터 파일의 비트맵 이미지에 메시지를 숨기거나, 오디오 파일에 데이터를 삽입하거나, 텍스트 파일에 메시지를 인코딩하여 은닉하는 등의 방법으로 구현된다. 이 기술은 기업의 민감한 정보를 불법적으로 유출하거나, 악성 코드를 숨겨 공격하는 방식과 같이 악의적으로 사용될 시 보안 및 개인 정보를 심각하게 위협하는 기술이 될 수 있다. 이에 따라 최근에는 스테가노그래피 탐지 기술이 중요한 요소로 대두되고 있다.

[0003] 스테가노그래피를 강인하게 탐지하기 위해서는 메시지가 삽입된 이미지인 스테고(Stego)와 삽입되지 않은 이미지인 커버(Cover) 사이의 미세한 차이를 식별해야 한다. 이를 위해 최근 연구에서는 딥러닝(특히 CNN 활용)을 활용한 스테가노그래피 탐지 기술이 주목받고 있다. 그러나 딥러닝 기반의 스테가노그래피 탐지기술의 경우 모델의 복잡성과 연산 비용이 큰 문제점이 있다.

발명의 내용

해결하려는 과제

[0004] 본 명세서는 전술한 문제점을 해결하기 위해, 이미지 전체를 입력으로 사용하여 얻은 특징이 아닌, 메시지가 삽입된 채널의 픽셀값 빈도를 계산한 픽셀 히스토그램을 사용함으로써, 모델의 복잡성과 연산비용 측면에서 효율적인 기계학습 기반의 스테가노그래피 탐지 방법을 제공하는 것을 목적으로 한다.

[0005] 본 발명이 이루고자 하는 기술적 과제들은 이상에서 언급한 기술적 과제들로 제한되지 않으며, 언급되지 않은 또 다른 기술적 과제들은 이하의 발명의 상세한 설명으로부터 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

[0006] 본 명세서의 일 실시예에 따른 스테가노그래피 탐지방법은, 이미지를 수신하는 단계; 상기 이미지에서 픽셀 히스토그램 벡터를 추출하는 단계; 상기 픽셀 히스토그램 벡터를 상기 이미지의 특징벡터로 정의하여 사전에 학습된 스테가노그래피 탐지모델에 입력하는 단계; 및 상기 스테가노그래피 탐지모델의 출력에 기초하여 상기 이미지를 커버(cover) 이미지와 스테고(stego) 이미지로 분류하는 단계;를 포함하고, 상기 스테가노그래피 탐지모델은, 다층퍼셉트론(MLP, Multi-Layer Perceptron) 기반의 네트워크이다.

[0007] 상기 스테가노그래피 탐지모델의 학습 데이터는, 상기 커버 이미지와 제1 레이블 데이터 쌍, 상기 스테고 이미지와 제2 레이블 데이터 쌍으로 구성될 수 있다.

[0008] 상기 스테고 이미지의 제2 픽셀 히스토그램에서 상기 커버 이미지의 제1 픽셀 히스토그램과 진동 폭에서 차이가 존재하는 범위의 데이터들이 상기 학습 데이터로 이용될 수 있다.

[0009] 상기 픽셀 히스토그램 벡터를 추출하는 단계는, 상기 픽셀 히스토그램의 값들 중 양 끝단의 값이 제거된 체 추출되는 값일 수 있다.

[0010] 상기 픽셀 히스토그램 벡터는 상기 이미지의 RGB 채널 중 적어도 하나의 채널의 픽셀 히스토그램 벡터일 수 있다.

[0011] 상기 픽셀 히스토그램 벡터는 상기 이미지의 RGB 채널 중 R 채널의 픽셀 히스토그램 벡터일 수 있다.

[0012] 본 명세서의 다른 실시예에 따른 스테가노그래피 이미지를 탐지하는 장치는, 적어도 하나의 프로그램이 기록된 메모리; 및 상기 프로그램을 실행하는 프로세서;를 포함하되, 상기 프로그램은, 수신된 이미지에서 픽셀 히스토그램 벡터를 추출하도록 하는 명령, 상기 픽셀 히스토그램 벡터를 상기 이미지의 특징벡터로 정의하여 사전에 학습된 스테가노그래피 탐지모델에 입력하도록 하는 명령을 포함하고, 상기 스테가노그래피 탐지모델은, 다층퍼셉트론(MLP, Multi-Layer Perceptron) 기반의 네트워크이며, 상기 스테가노그래피 탐지모델의 출력에 기초하여 상기 이미지는 커버(cover) 이미지와 스테고(stego) 이미지로 분류된다.

발명의 효과

[0013] 본 명세서의 일 실시예에 따르면, 이미지 데이터를 활용하는 종래의 방법에 비해 픽셀 히스토그램의 특징 벡터를 활용함으로써, 모델의 경량화, 추론 시간 및 정확도면에서 우수한 성능을 구현할 수 있다.

[0014] 본 발명에서 얻을 수 있는 효과는 이상에서 언급한 효과로 제한되지 않으며, 언급하지 않은 또 다른 효과들은 아래의 기재로부터 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

도면의 간단한 설명

[0015] 본 명세서에 관한 이해를 돕기 위해 상세한 설명의 일부로 포함되는, 첨부 도면은 본 명세서에 대한 실시예를 제공하고, 상세한 설명과 함께 본 명세서의 기술적 특징들을 설명한다.

도 1은 본 명세서의 일 실시예에 따른 스테가노그래피 탐지방법을 구현하기 위한 컴퓨팅 장치의 블록 구성도이다.

도 2는 본 명세서의 일 실시예에 적용되는 스테가노그래피 이미지를 설명하기 위한 도면이다.

도 3은 본 명세서의 일 실시예에 따른 스테가노그래피 탐지방법의 흐름도이다.

도 4는 본 명세서의 일 실시예에 따라 커버 이미지와 스테고 이미지의 픽셀값 분포의 차이를 설명하기 위한 도면이다.

도 5는 본 명세서의 일 실시예에 따른 스테가노그래피 탐지모델의 학습 구조를 설명하기 위한 도면이다.

본 명세서에 관한 이해를 돕기 위해 상세한 설명의 일부로 포함되는, 첨부 도면은 본 명세서에 대한 실시예를 제공하고, 상세한 설명과 함께 본 발명의 기술적 특징을 설명한다.

발명을 실시하기 위한 구체적인 내용

[0016] 이하, 첨부된 도면을 참조하여 본 명세서에 개시된 실시예를 상세히 설명하되, 도면 부호에 관계없이 동일하거나 유사한 구성요소는 동일한 참조 번호를 부여하고 이에 대한 중복되는 설명은 생략하기로 한다. 이하의 설명에서 사용되는 구성요소에 대한 접미사 "모듈" 및 "부"는 명세서 작성의 용이함만이 고려되어 부여되거나 혼용되는 것으로서, 그 자체로 서로 구별되는 의미 또는 역할을 갖는 것은 아니다. 또한, 본 명세서에 개시된 실시예를 설명함에 있어서 관련된 공지 기술에 대한 구체적인 설명이 본 명세서에 개시된 실시예의 요지를 흐릴 수 있다고 판단되는 경우 그 상세한 설명을 생략한다. 또한, 첨부된 도면은 본 명세서에 개시된 실시예를 쉽게 이해할 수 있도록 하기 위한 것일 뿐, 첨부된 도면에 의해 본 명세서에 개시된 기술적 사상이 제한되지 않으며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다.

[0017] 제1, 제2 등과 같이 서수를 포함하는 용어는 다양한 구성요소들을 설명하는데 사용될 수 있지만, 상기 구성요소들은 상기 용어들에 의해 한정되지는 않는다. 상기 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다.

[0018] 어떤 구성요소가 다른 구성요소에 "연결되어" 있다거나 "접속되어" 있다고 언급된 때에는, 그 다른 구성요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있지만, 중간에 다른 구성요소가 존재할 수도 있다고 이해되어야 할 것이다. 반면에, 어떤 구성요소가 다른 구성요소에 "직접 연결되어" 있다거나 "직접 접속되어" 있

다고 언급된 때에는, 중간에 다른 구성요소가 존재하지 않는 것으로 이해되어야 할 것이다.

[0019] 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다.

[0020] 본 출원에서, "포함한다" 또는 "가지다" 등의 용어는 명세서상에 기재된 특징, 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

[0021] 도 1은 본 명세서의 일 실시예에 따른 스테가노그래피 탐지방법을 구현하기 위한 컴퓨팅 장치의 블록 구성도이다.

[0022] 도 1을 참조하면, 스테가노그래피를 탐지하기 위한 장치(100)는, 컴퓨터 장치로서, 적어도 하나의 프로세서(110), 메모리(120), 입력부(130), 출력부(140), 저장부(150), 네트워크와 연결되어 통신을 수행하는 통신부(160) 및 데이터 학습부(170), 스테가노그래피 탐지부(180)를 포함할 수 있다. 상기 장치(100)에 포함된 각각의 구성 요소들은 버스(bus)(190)에 의해 연결되어 서로 통신을 수행할 수 있다.

[0023] 상기 컴퓨팅 장치는 AI 프로세싱 결과를 다른 기기에 제공하는 클라우드 환경의 디바이스일 수도 있다. 상기 컴퓨팅 장치(100)는 신경망을 학습할 수 있는 컴퓨팅 장치로서, 서버, 데스크탑 PC, 노트북 PC, 태블릿 PC 등과 같은 다양한 전자 장치로 구현될 수 있다. 또한, 상기 컴퓨팅 장치는 클라우드 환경의 서버와 데이터 통신을 수행하는 엣지 환경의 다비이스일 수도 있다.

[0024] 프로세서(110)는 메모리(120) 및/또는 저장부(150)에 저장된 프로그램을 이용하여 신경망을 학습할 수 있다. 특히, 프로세서(110)는 장치(100)의 동작을 위해 신경망을 학습할 수 있다. 여기서, 장치(100)의 동작을 위한 신경망 인간의 뇌 구조를 컴퓨터 상에서 모의하도록 설계될 수 있으며, 인간의 신경망의 뉴런(neuron)을 모의하는, 가중치를 갖는 복수의 네트워크 노드들을 포함할 수 있다. 복수의 네트워크 모드들은 뉴런이 시냅스(synapse)를 통해 신호를 주고 받는 뉴런의 시냅틱 활동을 모의하도록 각각 연결 관계에 따라 데이터를 주고 받을 수 있다. 여기서 신경망은 신경망 모델에서 발전한 딥러닝 모델을 포함할 수 있다. 딥러닝 모델에서 복수의 네트워크 노드들은 서로 다른 레이어에 위치하면서 연결 관계에 따라 데이터를 주고 받을 수 있다. 신경망 모델의 예는 심층 신경망(DNN, deep neural networks), 합성곱 신경망(CNN, convolutional deep neural networks), 순환 신경망(RNN, Recurrent Boltzmann Machine), 제한 볼츠만 머신(RBM, Restricted Boltzmann Machine), 심층 신뢰 신경망(DBN, deep belief networks), 심층 Q-네트워크(Deep Q-Network)와 같은 다양한 딥 러닝 기법들을 포함하며, 컴퓨터비젼, 음성인식, 자연어처리, 음성/신호처리 등의 분야에 적용될 수 있다.

[0025] 한편, 전술한 바와 같은 기능을 수행하는 프로세서는 범용 프로세서(예를 들어, CPU)일 수 있으나, 인공지능 학습을 위한 AI 전용 프로세서(예를 들어, GPU)일 수 있다.

[0026] 메모리(120) 및 저장부(150) 각각은 휘발성 저장 매체 및 비휘발성 저장 매체 중에서 적어도 하나로 구성될 수 있다. 예를 들어, 메모리(120)는 읽기 전용 메모리(read only memory, ROM) 및 랜덤 액세스 메모리(random access memory, RAM) 중에서 적어도 하나로 구성될 수 있다.

[0027] 메모리(120)는 장치(100)의 동작에 필요한 각종 프로그램 및 데이터를 저장할 수 있다. 메모리(120)는 비 휘발성 메모리, 휘발성 메모리, 플래시 메모리(flash-memory), 하드디스크 드라이브(HDD) 또는 솔리드 스테이트 드라이브(SDD) 등으로 구현할 수 있다. 메모리(120)는 프로세서(110)에 의해 액세스되며, 데이터의 독취/기록/수정/삭제/갱신 등이 수행될 수 있다. 또한, 메모리(120)는 본 명세서의 일 실시예에 따른 이미지 분류 및/또는 생성을 위한 AI 모델을 학습하기 위한 위한 학습 알고리즘을 통해 생성된 신경망 모델(예를 들어, 기계학습 모델 또는 딥 러닝 모델)을 저장할 수 있다.

[0028] 또한, 프로세서(110)는 메모리(120) 및 저장부(150) 중에서 적어도 하나에 저장된 프로그램 명령(program command)을 실행할 수 있다. 프로그램 명령은 수신된 이미지에서 픽셀 히스토그램 벡터를 추출하도록 하는 명령, 상기 픽셀 히스토그램 벡터를 상기 이미지의 특징벡터로 정의하여 사전에 학습된 스테가노그래피 탐지모델에 입력하도록 하는 명령을 포함하고, 상기 스테가노그래피 탐지모델은, 다층퍼셉트론(MLP, Multi-Layer Perceptron) 기반의 네트워크이며, 상기 스테가노그래피 탐지모델의 출력에 기초하여 상기 이미지는 커버(cover) 이미지와 스테고(stego) 이미지로 분류할 수 있다.

[0029] 한편, 프로세서(110)는 신경망 모델의 기능 예를 들어, 데이터 분류/인식을 위해 신경망을 학습하는 데이터 학습부(170)를 포함할 수 있다. 예를 들어, 데이터 학습부(170)는 데이터 분류/인식을 위하여 어떤 학습 데이터를

이용할지, 학습 데이터를 이용하여 데이터를 어떻게 분류하고 인식할지에 관한 기준을 학습할 수 있다. 데이터 학습부(170)는 학습에 이용될 학습 데이터를 획득하고, 획득된 학습데이터를 신경망 모델에 적용함으로써, 모델을 학습할 수 있다.

[0030] 데이터 학습부(170)는 적어도 하나의 하드웨어 칩 형태로 제작되어 장치(100)에 탑재될 수 있다. 예를 들어, 데이터 학습부(170)는 인공지능(AI)을 위한 전용 하드웨어 칩 형태로 제작될 수도 있고, 범용 프로세서(CPU) 또는 그래픽 전용 프로세서(GPU)의 일부로 제작되어 장치(100)에 탑재될 수도 있다. 또한, 데이터 학습부(170)는 소프트웨어 모듈로 구현될 수 있다. 소프트웨어 모듈(또는 인스트럭션(instruction)을 포함하는 프로그램 모듈)로 구현되는 경우, 소프트웨어 모듈은 컴퓨터로 읽을 수 있는 판독 가능한 비일시적 판독 가능 기록 매체(non-transitory computer readable media)에 저장될 수 있다. 이 경우, 적어도 하나의 소프트웨어 모듈은 OS(Operating System)에 의해 제공되거나, 애플리케이션에 의해 제공될 수 있다.

[0031] 데이터 학습부(170)는 학습 데이터 획득부(171) 및 모델 학습부(172)를 포함할 수 있다.

[0032] 학습 데이터 획득부(171)는 데이터를 분류하고 인식하기 위한 신경망 모델에 필요한 학습 데이터를 획득할 수 있다.

[0033] 모델 학습부(172)는 획득된 학습 데이터를 이용하여, 신경망 모델이 소정의 데이터를 어떻게 분류할지에 관한 판단 기준을 가지도록 학습할 수 있다. 이때 모델 학습부(172)는 학습 데이터 중 적어도 일부를 판단 기준으로 이용하는 지도 학습(supervised learning)을 통하여, 신경망 모델을 학습시킬 수 있다. 또는 모델 학습부(172)는 지도 없이 학습 데이터를 이용하여 스스로 학습함으로써, 판단 기준을 발견하는 비지도 학습(unsupervised learning)을 통해 신경망 모델을 학습시킬 수 있다. 또한, 모델 학습부(172)는 학습에 따른 상황 판단의 결과가 올바른지에 대한 피드백을 이용하여 강화 학습(reinforcement learning)을 통하여, 신경망 모델을 학습시킬 수 있다. 또한, 모델 학습부(172)는 오류 역전파법(error back-propagation) 또는 경사 하강법(gradient decent)을 포함하는 학습 알고리즘을 이용하여 신경망 모델을 학습시킬 수 있다.

[0034] 신경망 모델이 학습되면, 모델 학습부(172)는 학습된 신경망 모델을 메모리에 저장할 수 있다.

[0035] 데이터 학습부(170)는 인식 모델의 분석 결과를 향상시키거나, 인식 모델의 생성에 필요한 리소스 또는 시간을 절약하기 위해 학습 데이터 전처리부(미도시) 및 학습 데이터 선택부(미도시)를 더 포함할 수도 있다.

[0036] 학습 데이터 전처리부는 획득된 데이터가 상황 판단을 위한 학습에 이용될 수 있도록, 획득된 데이터를 전처리할 수 있다. 예를 들어, 학습 데이터 전처리부는, 모델 학습부(172)가 이미지 인식 및/또는 생성을 위한 학습을 위하여 획득된 학습 데이터를 이용할 수 있도록, 획득된 데이터를 기 설정된 포맷으로 가공할 수 있다.

[0037] 또한, 학습 데이터 선택부는, 학습 데이터 획득부(171)에서 획득된 학습 데이터 또는 전처리부에서 전처리된 학습 데이터 중 학습에 필요한 데이터를 선택할 수 있다.선택된 학습 데이터는 모델 학습부(172)에 제공될 수 있다.

[0038] 또한, 데이터 학습부(170)는 신경망 모델의 분석 결과를 향상시키기 위하여 모델 평가부(미도시)를 더 포함할 수도 있다.

[0039] 모델 평가부는, 신경망 모델에 평가 데이터를 입력하고, 평가 데이터로부터 출력되는 분석 결과가 소정 기준을 만족하지 못하는 경우, 모델 학습부(172)로 하여금 다시 학습하도록 할 수 있다. 이 경우, 평가 데이터는 인식 모델을 평가하기 위한 기 정의된 데이터일 수 있다. 일 예로, 모델 평가부는 평가 데이터에 대한 학습된 인식 모델의 분석 결과 중, 분석 결과가 정확하지 않은 평가 데이터의 개수 또는 비율이 미리 설정된 임계치를 초과하는 경우, 소정 기준을 만족하지 못한 것으로 평가할 수 있다.

[0040] 스테가노그래피 탐지부(180)는 스테고 이미지와 커버 이미지 간의 미세한 차이를 정밀하게 식별하기 위한 스테그어날리시스(steganalysis) 기능을 수행한다. 스테가노그래피 탐지부(180)는 데이터 학습부(170)를 학습된 스테가노그래피 탐지모델을 이용하여 커버 이미지와 스테고 이미지를 분류할 수 있다.

[0041] 도 2는 본 명세서의 일 실시예에 적용되는 스테가노그래피 이미지를 설명하기 위한 도면이다. 도 2를 참조하면, 커버 이미지에 스테가노그래피가 적용되어 암호(예를 들어, 메시지)가 삽입되는 구조를 나타낸다. 커버 이미지에 삽입되는 암호는 {-1,0,1}로 구성될 수 있으며, 본 명세서에서는 이미지의 RGB 채널 중 Red 채널에만 스테고 시그널을 삽입한 경우로 가정하여 설명한다. 도 2에서 커버 이미지와 스테고 이미지 각각에 대하여 각각 잔차(residual) 이미지를 생성하여 확인하는 경우 R 채널 이미지에서 미세하게 차이가 존재하는 것을 확인할 수 있지만, 도 2에 도시된 커버 이미지와 스테고 이미지만으로는 실질적인 차이를 확인할 수 없는 바, 스테가노그래

피 탐지기술이 필요할 수 있다.

[0042] 도 3은 본 명세서의 일 실시예에 따른 스테가노그래피 탐지방법의 흐름도이다. 도 3에 도시된 스테가노그래피 탐지방법은 도 1의 컴퓨팅 장치의 프로세서(110)의 제어를 통해 구현될 수 있다.

[0043] 도 3을 참조하면, 프로세서(110)는 이미지를 수신한다(S300). 상기 이미지는 스테가노그래피 처리가 된 스테고 이미지일 수도 있고, 커버 이미지일 수도 있다. 프로세서(110)는 상기 이미지에서 픽셀 히스토그램 벡터를 추출할 수 있다(S310).

[0044] 프로세서(110)는 상기 픽셀 히스토그램 벡터를 이미지의 특징벡터로 정의하고, 사전에 학습된 스테가노그래피 탐지모델에 입력할 수 있다(S320). 그리고 프로세서(110)는 스테가노그래피 탐지모델의 출력에 기초하여 이미지를 커버 이미지와 스테고 이미지로 분류할 수 있다(S330).

[0045] 본 명세서의 일 실시예에 따라 이미지에서 픽셀 히스토그램은 아래 수학식을 참조할 수 있다.

수학식 1

[0046] $$h_i(n)=\sum_{j=1}^{d}\sum_{k=1}^{d}\delta\left(p_i^{(i,k)},n\right)$$

수학식 2

[0047] $$\delta(x,y)=\begin{cases}1 \text{ if } x=y,\\ 0\ otherwise.\end{cases}$$

[0048] 여기서 $h_i(n)$는 d X d크기를 갖는 이미지 v_i의 R 채널 픽셀값 n의 빈도수 즉, R 채널의 픽셀 히스토그램을 의미한다. $p_i^{(j,k)}$ 는 v_i 의 j행 k열에 해당하는 R 채널 픽셀값을 의미하며, {0,1,쩜,255} 중 하나의 값을 가진다. 델타함수 δ 는 수학식 2와 같이 정의되며 이 함수는 i와 j가 같을 때 1을, 그렇지 않을 때는 0을 반환한다.

[0049] 본 명세서의 일 실시예에 따른 스테가노그래피 탐지모델의 학습에 이용되는 입력 데이터로는 수학식 3과 같이 정의되는 이미지 v_i 의 픽셀 히스토그램 데이터 h_i 를 이용할 수 있다.

수학식 3

[0050] $$\boldsymbol{h}_i=\{h_i(n)\}_{n=0}^{255}$$

[0051] 도 4는 본 명세서의 일 실시예에 따라 커버 이미지와 스테고 이미지의 픽셀값 분포의 차이를 설명하기 위한 도면이다. 도 4를 참조하면, (a)는 커버 이미지 $v_{i,c}$ 와 대응하는 픽셀 히스토그램 $h_{i,c}$ 를 시각화한 것이며, (b)는 스테고(stego) 이미지 $v_{i,s}$ 와 대응하는 픽셀 히스토그램 $h_{i,s}$ 시각화한 예시이다. 두 이미지의 픽셀 히스토그램 사이에는 명확한 차이가 보인다. 즉, 스테고 이미지의 픽셀 히스토그램 $h_{i,s}$은 커버 이미지의 픽셀 히스토그램 $h_{i,c}$ 에 비해 더 넓은 진동 폭을 가진다. 이는 스테가노그래피로 삽입된 메시지로 인한 픽셀 값의 변화를 의미하며, 숨겨진 정보의 나타낼 수 있다.

[0052] 즉, 이러한 변화는 스테가노그래피 과정에서 정보가 삽입됨에 따라 이미지의 픽셀 값 분포에 영향을 미치는 것으로 해석된다. 특히 히스토그램의 50<i<200 범위에서 스테가노그래피가 가장 활발하게 적용된 것임을 확인할 수 있다.

[0053] 이 범위에서 주요 변화는 원본 이미지의 정보를 가장 잘 표현하고, 동시에 스테가노그래피 신호를 가장 잘 숨기는 영역이 이 범위에 있음을 시사한다.

[0054] 도 5는 본 명세서의 일 실시예에 따른 스테가노그래피 탐지모델의 학습 구조를 설명하기 위한 도면이다.

[0055] 본 명세서의 일 실시예에 따른 스테가노그래피 탐지모델의 학습 데이터는 커버 이미지와 제1 레이블 데이터의 쌍과, 스테고 이미지와 제2 레이블 데이터의 쌍으로 구성될 수 있다. 상기 제1 레이블 데이터 및 제2 레이블 데이터는 정답 데이터를 의미할 수 있다. 분류 클래스는 커버 이미지와 스테고 이미지로 각각 커버 이미지와 커버 이미지에 메시지가 삽입된 스테고 이미지에 대한 이진 분류 과정을 의미할 수 있다. 일 실시예에 따라 학습 데이터 셋 $D = \{v_i, y_i\}_{i=0}^{20000}$ 는 i 번째 이미지 데이터 샘플 v_i 와, 이에 대응되는 정답 클래스 $y_i \in \{0, 1\}$ (0: cover, 1: stego)의 쌍으로 구성되어 있으며, 각 클래스에 해당하는 이미지 수는 각각 10,000개로 동일한 비율을 가질 수 있다.

[0056] 도 5를 참조하면, 프로세서(110)는 먼저 이미지 v_i 에서 픽셀 히스토그램 벡터 h_i 를 추출한 뒤, h_i 자체를 이미지의 특징벡터로 활용하여 다층퍼셉트론(MLP) 기반의 네트워크의 입력으로 사용한다.

[0057] MLP 네트워크는 가중치 (W_α, W_β)와 편향 (b_α, b_β)로 이루어지는 두 개의 선형 계층으로 구성되며, 각 레이어는 수학식 4로 표현되는 LeakyReLU 활성화 함수를 포함한다. 또한, 수학식 5의 손실함수는 Binary Cross Entropy(BCE) 손실함수를 사용한다. 수학식 5에서 y_i 는 0 또는 1의 값을 갖는 정답 클래스를 의미하고, $\hat{y}_i$ 는 모델의 예측 값을 의미한다. N은 학습 데이터의 수를 나타낸다.

수학식 4

[0058] $$\mathrm{LeakyReLU}(x) = \max(0.01 \times x, x)$$

수학식 5

[0059] $$L(y, \hat{y}) = -\frac{1}{N}\sum_{i}^{N}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)]$$

[0060] $$\hat{y}_i = W_\beta \cdot LeakyReLU(W_\alpha \cdot \boldsymbol{h}_i + b_\alpha) + b_\beta$$

[0061] $L(y, \hat{y})$ 은 실제 정답 클래스인 y 와 예측 값 $\hat{y}$ 사이의 차이를 최소화하는 방향으로 네트워크가 학습을 진행하게 된다. 본 명세서의 일 실시예에 따른 스테가노그래피 탐지모델의 학습방법의 중요한 효과 중 하나는 이미지의 특징을 복잡한 컨볼루션 신경망 기반의 특징 추출기를 사용하지 않고, 픽셀 히스토그램을 통해 추출하기 때문에 두 개의 선형 레이어 만을 이용한 간단한 모델 구조임에도 불구하고, 원본 이미지를 직접 입력 데이터로 사용한 경우에 비해 우수한 성능을 보인다. 이는 픽셀 히스토그램이 이미지의 핵심적인 정보를 효율적으로 추출하기 때문이다.

[0062] 본 명세서는 이미지의 픽셀 히스토그램만을 이용하기 때문에 종래의 스테가노그래피 알고리즘보다 계산 복잡도가 크게 감소하는 이점이 있다. 입력 데이터의 차원을 비교할 때 예를 들어, 종래의 컨볼루션 신경망 기반의 방

법은 256X256 크기의 이미지를 입력으로 사용하므로 데이터의 차원이 65,536차원이된다. 하지만, 본 명세서의 일 실시예에 따른 스테가노그래피 탐지모델의 경우 256차원으로 줄어든다. 픽셀 히스토그램을 통해 이미지의 전반적인 통계적 특성을 추출할 수 있기 때문에 미세한 차이가 있는 스테고 이미지를 효과적으로 식별하는데 유리하다.

[0063] 본 명세서의 일 실시예에 따른 스테가노그래피 탐지모델 학습은 2개의 레이어로 구성된 MLP 기반의 네트워크를 사용하였으며, 첫 번째 레이어에서 LeakyReLU 활성화 함수를 적용하여 비선형성을 부여하였다. 또한, 과적합을 방지하기 위해 동일한 레이어에 0.1의 확률의 Dropout 기법을 적용하였다. 또한, Optimizer는 AdamW를 사용하였으며, Learning rate는 0.001, Leaning Rate Scheduler로는 MultiStepLR을 적용하였으며, Epoch는 15, 학습 배치 사이즈는 16으로 설정하여 학습하였다.

표 1

[0064]

Model	#Parameter	Latency Time	Acc.
Ours	51K	0.313 ms	99.87%
EfficinetNet-B2	9.262K	2.093 ms	98.65%

[0065] 표 1은 본 명세서에서 개시하는 스테가노그래피 탐지방법과 기존의 접근법인 CNN 기반 이미지 특징 추출기로 EfficientNet-B2를 사용하는 방법 사이의 성능 비교 결 과이다. 본 명세서의 일 실시예에 다른 스테가노그래피 탐지방법은, 적은 모델 파라미터와 추론 시간을 크게 줄이면서도, 더 높은 정확도를 달성한다. 모델 파라미터의 수는 EfficientNet-B2의 경우 약 9,262K 개인 반면, 제안하는 기법에서는 약 51K개로 약 181배 감소하였다. 또한 추론 시간의 경우 제안하는 기법이 0.0313ms, EfficientNet-B2 기반 모델은 2.096ms로 약 7배 단축되었다. 이는 추론에 소요되는 시간을 크게 줄여 시간 측면에서 효율성을 높일 수 있다는 결과를 나타낸다. 정확도 면에서도 제안하는 기법에서는 99.87%로 우수한 성능을 보였다. 이는 이미지 데이터를 픽셀 히스토그램 데이터로 변환하여 사용함으로써 더 효과적인 스테가노그래피 탐지가 가능하다는 것을 보여준다. 결과적으로, 본 명세서의 일 실시예에 다른 픽셀 히스토그램을 이용한 스테가노그래피 탐지방법은, 픽셀 히스토그램을 특징 벡터로 활용함으로써, 스테가노그래피 탐지모델은 경량화 되고, 추론 시간 및 정확도 면에서 기존 방식에 비해 효율인 방법인을 알 수 있다.

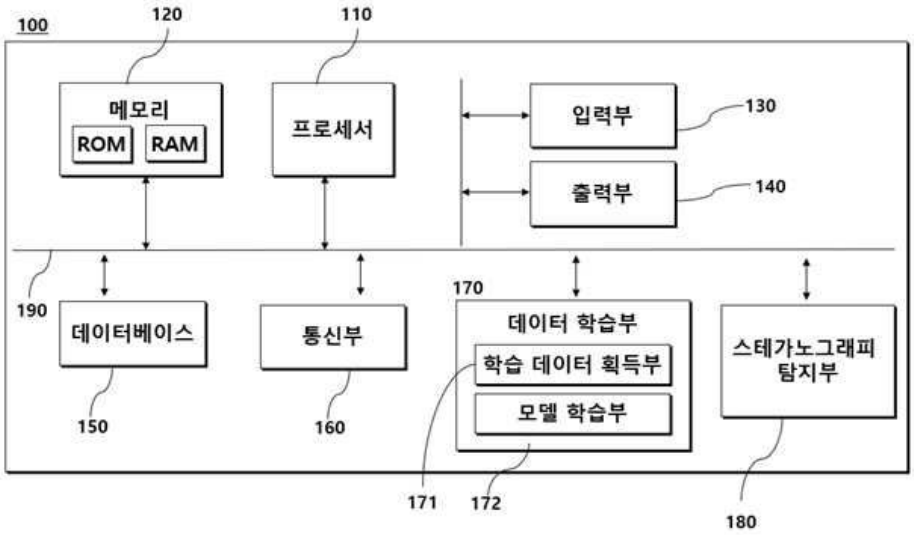
[0066] 한편, 본 명세서의 일 실시예에 따르면 프로세서(110)는 픽셀 히스토그램의 값들 중 양 끝단의 값(예를 들어, 0,255)의 경우, 극단적인 큰 값이 나타나는 경우가 있으며 이에 따라 양 끝단의 값을 제외하고 학습을 진행하는 경우 스테가노그래피 탐지의 안정성을 향상시킬 수 있다. 이에 따라 본 명세서의 일 실시예에 따라, 이미지에서 픽셀 히스토그램 벡터를 추출하는 과정에서 픽셀 히스토그램의 값을 중 양 끝단의 값을 제외할 수 있다.

[0067] 전술한 본 발명은, 프로그램이 기록된 매체에 컴퓨터가 읽을 수 있는 코드로서 구현하는 것이 가능하다. 컴퓨터가 읽을 수 있는 매체는, 컴퓨터 시스템에 의하여 읽힐 수 있는 데이터가 저장되는 모든 종류의 기록장치를 포함한다. 컴퓨터가 읽을 수 있는 매체의 예로는, HDD(Hard Disk Drive), SSD(Solid State Disk), SDD(Silicon Disk Drive), ROM, RAM, CD-ROM, 자기 테이프, 플로피 디스크, 광 데이터 저장 장치 등이 있으며, 또한 캐리어 웨이브(예를 들어, 인터넷을 통한 전송)의 형태로 구현되는 것도 포함한다. 따라서, 상기의 상세한 설명은 모든 면에서 제한적으로 해석되어서는 아니되고 예시적인 것으로 고려되어야 한다. 본 발명의 범위는 첨부된 청구항의 합리적 해석에 의해 결정되어야 하고, 본 발명의 등가적 범위 내에서의 모든 변경은 본 발명의 범위에 포함된다.

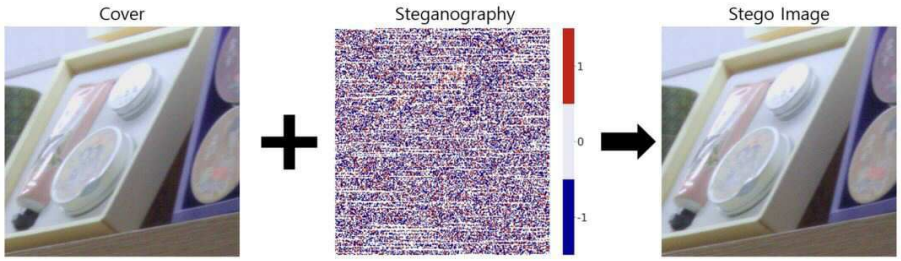
부호의 설명

도면

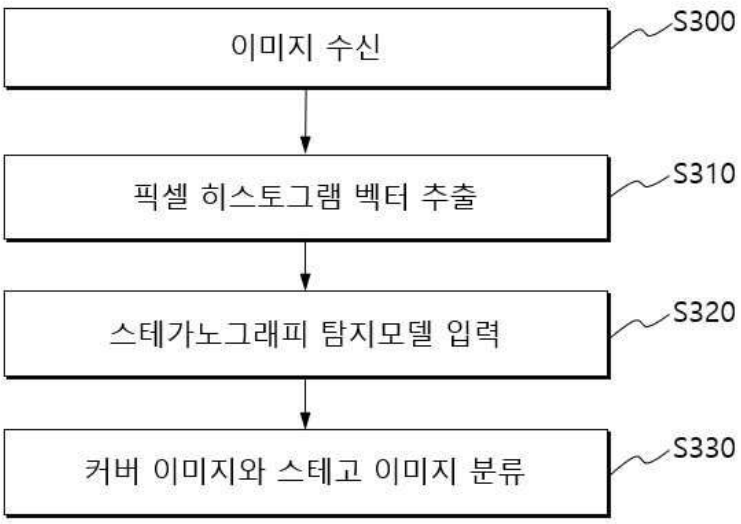
도면1



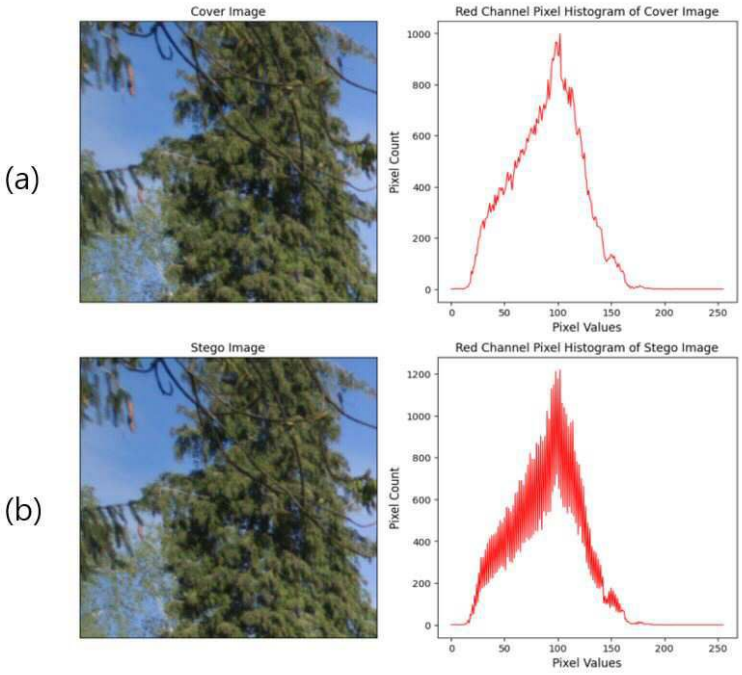
도면2



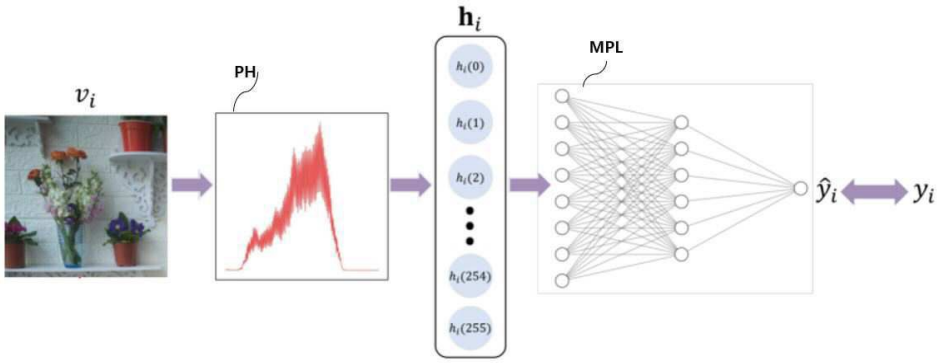
도면3



도면4


(a)
Cover Image
Red Channel Pixel Histogram of Cover Image
Pixel Count
Pixel Values
(b)
Stego Image
Red Channel Pixel Histogram of Stego Image
Pixel Count
Pixel Values

도면5


v_i
PH
$\mathbf{h}_i$
$h_i(0)$
$h_i(1)$
$h_i(2)$
$h_i(254)$
$h_i(255)$
MPL
$\hat{y}_i$
y_i

임시명세서(첨부)



아이콘을 더블 클릭하시면 임시명세서 파일이 열립니다.

본 공보 PDF는 첨부파일을 가지고 있습니다. Acrobat Reader PDF뷰어를 제공하지 않는 브라우저(크롬, 파이어폭스, 사파리 등)의 경우 첨부파일 열기가 제한되어 있으므로 Acrobat Reader PDF뷰어 설치 후 공보 PDF를 다운로드 받아 해당 뷰어에서 조회해주시기 바랍니다.