



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0054681
(43) 공개일자 2024년04월26일

- | | |
|--|--|
| <p>(51) 국제특허분류(Int. Cl.)
H04N 21/44 (2024.01) G06N 20/20 (2019.01)
G06N 3/08 (2023.01) G06V 40/16 (2022.01)
H04N 21/439 (2011.01)</p> <p>(52) CPC특허분류
H04N 21/44008 (2013.01)
G06N 20/20 (2021.08)</p> <p>(21) 출원번호 10-2022-0134934
(22) 출원일자 2022년10월19일
심사청구일자 2022년10월19일</p> | <p>(71) 출원인
울산과학기술원
울산광역시 울주군 언양읍 유니스트길 50</p> <p>(72) 발명자
백승렬
울산광역시 울주군 언양읍 유니스트길 50
정윤영
울산광역시 울주군 언양읍 유니스트길 50</p> <p>(74) 대리인
전용준</p> |
|--|--|

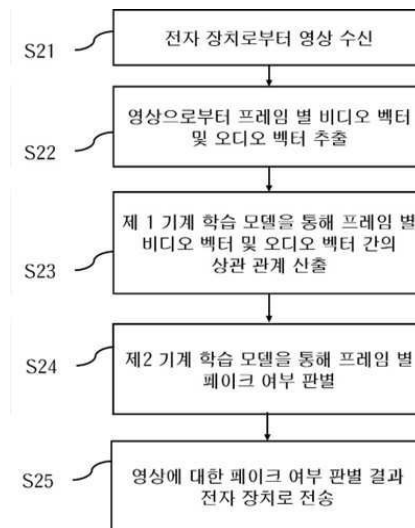
전체 청구항 수 : 총 15 항

(54) 발명의 명칭 **페이크 영상 검증 방법 및 그 시스템**

(57) 요약

페이크 비디오 검증 방법 및 그 시스템을 개시한다. 본 방법은 동영상으로부터 프레임 별 비디오 벡터 및 오디오 벡터를 추출하고, 프레임 별로 추출된 비디오 벡터 및 오디오 벡터 간의 상관 관계 값을 산출하고, 프레임 별로 산출된 상관 관계 값에 기초하여 동영상이 페이크 비디오인지 여부를 판별한다.

대표도 - 도2



(52) CPC특허분류

G06N 3/08 (2023.01)

G06V 40/16 (2022.01)

H04N 21/4394 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711159656
과제번호	2021-0-01778-002
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	차세대인공지능핵심원천기술개발
연구과제명	실제와 지각 역치 이하 수준까지 동일한 특성을 갖는 인물 영상 합성 기술 및 판별
기술 개발	
기 여 율	1/1
과제수행기관명	울산과학기술원
연구기간	2022.02.01 ~ 2023.01.31

명세서

청구범위

청구항 1

전자 장치가 페이크 비디오를 검증하는 방법에 있어서,
동영상으로부터 프레임 별 비디오 벡터 및 오디오 벡터를 추출하는 단계;
상기 프레임 별로 상기 추출된 비디오 벡터 및 오디오 벡터 간의 상관 관계 값을 산출하는 단계; 및
상기 프레임 별로 산출된 상관 관계 값에 기초하여 상기 동영상의 페이크 비디오인지 여부를 판별하는 단계를 포함하는, 방법.

청구항 2

제 1 항에 있어서,
상기 동영상의 페이크 비디오인지 여부를 판별하는 단계는,
상기 프레임 별 상관 관계 값과 기 설정된 기준 값을 비교한 결과에 따라 상기 비디오 벡터 및 오디오 벡터 간의 관계가 페이크인지 여부를 판단하는 단계; 및
상기 동영상에 대해 프레임 별로 페이크 여부를 판단한 결과를 출력하는 단계를 포함하는, 방법.

청구항 3

제 1 항에 있어서,
상기 동영상으로부터 프레임 별 비디오 벡터 및 오디오 벡터를 추출하는 단계는,
상기 동영상에 포함된 이미지 데이터를 기 학습된 3DMM(3D Morphable face Model) 기반의 신경망에 입력하여 비디오 벡터로 변환하는 단계; 및
상기 동영상에 포함된 오디오 데이터를 Mel-frequency Cepstral Coefficients (MFCC) 모델을 통해 오디오 벡터로 변환하는 단계를 포함하는, 방법.

청구항 4

제 1 항에 있어서,
상기 추출된 비디오 벡터 및 오디오 벡터 간의 상관 관계 값을 산출하는 단계는,
상기 비디오 벡터 및 오디오 벡터를 결합하여 인공 신경망의 FC 레이어(Fully Connected Layer)에 입력하는 단계; 및
상기 FC 레이어를 통해 0과 1사이의 실수 값 중 각 프레임의 비디오 벡터 및 오디오 벡터 간의 상관 관계에 대응하는 실수 값을 출력하는 단계를 포함하는, 방법.

청구항 5

제 1 항에 있어서,
상기 동영상의 페이크 비디오인지 여부를 판별하는 단계는,
상기 프레임 별 상관 관계 값을 결합한 벡터를 인공 신경망의 FC 레이어(Fully Connected Layer)에 입력하는 단계; 및
상기 FC 레이어를 통해 상기 동영상에 대한 비디오 벡터 및 오디오 벡터 간의 상관 관계가 페이크인지 여부를 나타내는 0 또는 1의 값을 출력하는 단계를 포함하는, 방법.

청구항 6

제 5항에 있어서,

상기 동영상상이 페이크 비디오인지 여부를 판별하는 단계는,

상기 동영상에 포함된 복수의 프레임들의 상기 상관 관계 값을 순차적으로 입력 받으며, 기 설정된 프레임 개수 단위로 상기 상관 관계 값을 결합한 벡터에 대한 페이크 비디오 여부를 판별하는, 방법.

청구항 7

제 6 항에 있어서,

상기 동영상상이 페이크 비디오인지 여부를 판별하는 단계는,

상기 상관 관계 값을 결합한 벡터에 대해 최초로 비디오 벡터 및 오디오 벡터 간의 관계가 페이크로 판별된 경우, 상기 동영상을 페이크 비디오로써 판별하는 단계를 포함하는, 방법.

청구항 8

적어도 하나의 클라이언트 장치와 통신하는 서버가 페이크 비디오를 검증하는 방법에 있어서,

상기 적어도 하나의 클라이언트 장치로부터 동영상을 수신하는 단계;

동영상으로부터 추출된 프레임 별 비디오 벡터 및 오디오 벡터 간의 상관 관계 값을 산출하는 단계;

상기 프레임 별로 산출된 상관 관계 값에 기초하여 상기 동영상상이 페이크 비디오인지 여부를 판별하는 단계; 및

상기 판별된 결과를 상기 클라이언트 장치로 전송하는 단계를 포함하는 방법.

청구항 9

제 8 항에 있어서,

상기 동영상상이 페이크 비디오인지 여부를 판별하는 단계는,

상기 프레임 별 상관 관계 값과 기 설정된 기준 값을 비교한 결과에 따라 상기 비디오 벡터 및 오디오 벡터 간의 관계가 페이크인지 여부를 판단하는 단계; 및

상기 동영상에 대해 프레임 별로 페이크 여부를 판단한 결과를 출력하는 단계를 포함하는, 방법.

청구항 10

제 8 항에 있어서,

상기 동영상상이 페이크 비디오인지 여부를 판별하는 단계는,

상기 프레임 별 상관 관계 값과 기 설정된 기준 값을 비교한 결과에 따라 상기 비디오 벡터 및 오디오 벡터 간의 관계가 페이크인지 여부를 판단하는 단계; 및

상기 동영상에 대해 프레임 별로 페이크 여부를 판단한 결과를 출력하는 단계를 포함하는, 방법.

청구항 11

제8항에 있어서,

상기 동영상상이 페이크 비디오인지 여부를 판별하는 단계는,

상기 동영상에 포함된 복수의 프레임들의 상기 상관 관계 값을 순차적으로 입력 받으며, 기 설정된 프레임 개수 단위로 상기 상관 관계 값을 결합한 벡터에 대한 페이크 비디오 여부를 판별하는, 방법.

청구항 12

제 11 항에 있어서,

상기 동영상상이 페이크 비디오인지 여부를 판별하는 단계는,

상기 상관 관계 값을 결합한 벡터에 대해 최초로 비디오 벡터 및 오디오 벡터 간의 관계가 페이크로 판별된 경우, 상기 동영상상을 페이크 비디오로써 판별하는 단계를 포함하는, 방법.

청구항 13

제 8 항에 있어서,

상기 클라이언트 장치로부터 동영상상을 수신하는 단계는,

상기 클라이언트 장치에 의해 동영상상에 대해 프레임 별로 비디오 벡터 및 오디오 벡터를 추출한 데이터를 수신하는 것인 방법.

청구항 14

하나 이상의 인스트럭션을 저장하는 메모리; 및

상기 메모리에 저장된 상기 하나 이상의 인스트럭션을 실행하는 프로세서를 포함하고,

상기 프로세서는 상기 하나 이상의 인스트럭션을 실행함으로써,

제1 항의 방법을 수행하는, 전자 장치.

청구항 15

하드웨어인 컴퓨터와 결합되어, 제1 항의 방법을 수행할 수 있도록 컴퓨터에서 독출가능한 기록매체에 저장된 컴퓨터프로그램.

발명의 설명

기술 분야

[0001] 본 발명은 페이크 영상을 검증하기 위한 방법 및 이를 구현하기 위한 시스템에 관한 것이다.

배경 기술

[0002] 전 세계적으로 동영상 플랫폼의 사용이 증가하였으며, 특히 개인 동영상 서비스를 이용하는 수요가 폭발적으로 증가함에 따라 개인의 안전 및 권리 확보를 위해 영상에 대한 진위 여부를 검증하는 것이 매우 중요한 이슈가 되었다. 이러한 개인 동영상 서비스뿐 아니라 사회적 정보를 제공하는 공공 시스템 및 방송 등에서도 동영상의 신뢰성은 가장 중요한 요소이다.

[0003] 그러나 최근 들어 딥 페이크 기술 분야가 발전함에 따라 원본 영상을 교묘하게 편집하여 위 변조된 영상 즉, 페이크 영상을 오용 또는 남용하는 사례 또한 크게 증가하였다.

[0004] 딥 페이크 기술이란 복수의 인물의 이미지 또는 비디오를 서로 합성하여 진짜 같은 새로운 영상 데이터를 만들어 내는 기술로써, 인공지능 기술이 딥 페이크 기술에 적용됨에 따라 육안으로는 실제 존재하는 비디오와 페이크 비디오를 사실상 구별하기가 더욱 어려워지고 있다.

[0005] 특히, 원본 영상을 확인하기 어려운 상황이거나 정밀도가 매우 높은 페이크 영상의 경우 기존의 영상 검증 방법 만으로는 페이크 영상에 대한 진위 여부를 정확하고 신속하게 검증하기가 매우 어려운 실정이다.

[0006] 따라서, 별도의 원본 영상이 없더라도 페이크 영상 자체에 대해 정확한 검증 결과를 신속하게 얻을 수 있는 검증 기술의 개발이 필요하다.

선행기술문헌

특허문헌

[0007] (특허문헌 0001) 대한민국 등록 특허 공보 제10-2286468 호 (발명의 명칭: 딥 러닝을 통해 생성된 가짜 동영상의 무결성 검증 방법 및 시스템)

발명의 내용

해결하려는 과제

- [0008] 본 발명은 상술한 문제점을 해결하기 위한 것으로, 본 발명의 일 실시예는 비디오 데이터만으로는 검증하기 어려운 페이크 영상을 검증할 수 있는 방법 및 이를 구현하기 위한 시스템을 제공하고자 한다.
- [0009] 다만, 본 실시예가 이루고자 하는 기술적 과제는 상기된 바와 같은 기술적 과제들로 한정되지 않으며, 또 다른 기술적 과제들이 더 존재할 수 있다.

과제의 해결 수단

- [0010] 상술한 기술적 과제를 달성하기 위한 기술적 수단으로서, 본 발명의 일 측면에 따른 페이크 비디오를 검증하는 방법은, 동영상으로부터 프레임 별 비디오 벡터 및 오디오 벡터를 추출하는 단계; 상기 프레임 별로 상기 추출된 비디오 벡터 및 오디오 벡터 간의 상관 관계 값을 산출하는 단계; 및 상기 프레임 별로 산출된 상관 관계 값에 기초하여 상기 동영상이 페이크 비디오인지 여부를 판별하는 단계를 포함한다.
- [0011] 본 발명의 기타 구체적인 사항들은 상세한 설명 및 도면들에 포함되어 있다.

발명의 효과

- [0012] 본 발명의 일 실시예에 따르면, 영상의 비디오 및 오디오의 상관 관계(즉, 동기화 여부)를 검출함에 따라 비디오와 오디오 간의 불일치성을 감지할 수 있어, 페이크 영상 자체만으로도 진위 여부를 정확하게 판별할 수 있는 효과가 있다.
- [0013] 또한, 본 발명의 다른 실시예에 따르면, 영상의 프레임 단위로 비디오 및 오디오의 상관 관계 정도를 검출하여 페이크 여부를 판단함으로써, 전체 영상에 대해 검증하지 않고도 적은 양의 데이터 처리를 통해 신속하게 페이크 영상 여부를 검증할 수 있다.
- [0014] 본 발명의 효과들은 이상에서 언급된 효과로 제한되지 않으며, 언급되지 않은 또 다른 효과들은 아래의 기재로부터 통상의 기술자에게 명확하게 이해될 수 있을 것이다.

도면의 간단한 설명

- [0015] 도 1은 본 발명의 일 실시예에 따른 페이크 영상 검증 시스템을 도시한 도면이다.
- 도 2는 본 발명의 일 실시예에 따른 페이크 영상 검증 시스템의 동작 방법을 도시한 순서도이다.
- 도 3은 본 발명의 일 실시예에 따른 페이크 영상 검증 프로세스를 나타낸 도식도이다.
- 도 4는 본 발명의 일 실시예에 따른 오디오 벡터 추출 단계에서 사용되는 스펙트로그램의 일례를 나타낸 도면이다.
- 도 5는 본 발명의 일 실시예에 따른 비디오 벡터 및 오디오 벡터 간의 상관 관계 산출 프로세스를 나타낸 도식도이다.
- 도 6은 본 발명의 일 실시예에 따라 도 2 내지 도 5의 동작을 서버가 아닌 전자 장치가 수행하는 페이크 영상 검증 시스템을 도시한 일 예이다.
- 도 7은 본 발명의 일 실시 예에 따른 서버의 구성도이다.

발명을 실시하기 위한 구체적인 내용

- [0016] 본 발명의 이점 및 특징, 그리고 그것들을 달성하는 방법은 첨부되는 도면과 함께 상세하게 후술되어 있는 실시예들을 참조하면 명확해질 것이다. 그러나, 본 발명은 이하에서 개시되는 실시예들에 제한되는 것이 아니라 서로 다른 다양한 형태로 구현될 수 있으며, 단지 본 실시예들은 본 발명의 개시가 완전하도록 하고, 본 발명이 속하는 기술 분야의 통상의 기술자에게 본 발명의 범주를 완전하게 알려주기 위해 제공되는 것이며, 본 발명은 청구항의 범주에 의해 정의될 뿐이다.
- [0017] 본 명세서에서 사용된 용어는 실시예들을 설명하기 위한 것이며 본 발명을 제한하고자 하는 것은 아니다. 본 명세서에서, 단수형은 문구에서 특별히 언급하지 않는 한 복수형도 포함한다. 명세서에서 사용되는 "포함한다

(comprises)" 및/또는 "포함하는(comprising)"은 언급된 구성요소 외에 하나 이상의 다른 구성요소의 존재 또는 추가를 배제하지 않는다. 명세서 전체에 걸쳐 동일한 도면 부호는 동일한 구성 요소를 지칭하며, "및/또는"은 언급된 구성요소들의 각각 및 하나 이상의 모든 조합을 포함한다. 비록 "제1", "제2" 등이 다양한 구성요소들을 서술하기 위해서 사용되나, 이들 구성요소들은 이들 용어에 의해 제한되지 않음은 물론이다. 이들 용어들은 단지 하나의 구성요소를 다른 구성요소와 구별하기 위하여 사용하는 것이다. 따라서, 이하에서 언급되는 제1 구성요소는 본 발명의 기술적 사상 내에서 제2 구성요소일 수도 있음은 물론이다.

- [0018] 다른 정의가 없다면, 본 명세서에서 사용되는 모든 용어(기술 및 과학적 용어를 포함)는 본 발명이 속하는 기술 분야의 통상의 기술자에게 공통적으로 이해될 수 있는 의미로 사용될 수 있을 것이다. 또한, 일반적으로 사용되는 사전에 정의되어 있는 용어들은 명백하게 특별히 정의되어 있지 않는 한 이상적으로 또는 과도하게 해석되지 않는다.
- [0019] 본 명세서에서, 전자 장치는 적어도 하나의 프로세서를 포함하는 무선 통신 기능을 포함하는 하드웨어 장치를 의미하는 것이고, 실시 예에 따라 해당 하드웨어 장치에서 동작하는 소프트웨어적 구성도 포괄하는 의미로서 이해될 수 있다. 예를 들어, 전자 장치는 스마트폰, 휴대폰, 핸드 헬드 장치(hand-held device), 태블릿 PC, 데스크톱, 노트북 및 각 장치에서 구동되는 사용자 클라이언트 및 애플리케이션을 모두 포함하는 의미로서 이해될 수 있으며, 또한 이에 제한되는 것은 아니다.
- [0020] 이하, 첨부된 도면을 참조하여 본 발명의 실시예를 상세하게 설명한다.
- [0021] 본 명세서에서 설명되는 각 단계들은 단말에 의하여 수행되는 것으로 설명되나, 각 단계의 주체는 이에 제한되는 것은 아니며, 실시 예에 따라 각 단계들의 적어도 일부가 서로 다른 장치에서 수행될 수도 있다.
- [0022] 도 1은 본 발명의 일 실시예에 따른 페이크 영상 검증 시스템(1)을 도시한 도면이다.
- [0023] 도 2는 본 발명의 일 실시예에 따른 페이크 영상 검증 시스템(1)의 동작 방법을 도시한 순서도이고, 도 3은 본 발명의 일 실시예에 따른 페이크 영상 검증 프로세스를 나타낸 도식도이다.
- [0024] 본 발명의 일 실시예에 따른 페이크 영상 검증 시스템(1)은 서버(10), 복수의 전자 장치(20)를 포함한다. 서버(10)와 적어도 하나의 전자 장치(20) 및 각각의 전자 장치(20)는 서로 네트워크를 통해 통신할 수 있으며, 이때 각각의 전자 장치(20)는 클라이언트 장치(client device)로 동작할 수 있다.
- [0025] 서버(10)는 적어도 하나의 전자 장치(20)로부터 영상(즉, 복수의 비디오 프레임(frame))을 수신할 수 있다(S21).
- [0026] 예를 들어, 도 3에 도시된 바와 같이 수신된 영상은 적어도 하나의 인물이 포함된 이미지 데이터 및 음성과 같은 오디오 데이터를 포함한다.
- [0027] 서버(10)는 수신된 영상으로부터 프레임 별 비디오 벡터 및 오디오 벡터를 추출한다(S22).
- [0028] 구체적으로, 도 3에 도시된 바와 같이, 서버(10)는 비디오 데이터를 3DMM (3D Morphable face Model) 기반의 신경망에 입력하여 복수의 파라미터(parameter) 즉, 비디오 벡터(Fv)로 변환할 수 있다.
- [0029] 3DMM 은 3D 스캐너나 깊이 카메라를 이용하여 얼굴의 3차원 정보를 직접적으로 이용하여 3차원 형태와 질감 간의 통계적 모델을 표현한 것으로, 이미지 상에서 사람의 얼굴을 찾고 이를 표현하는 소수의 변수들을 벡터로 출력한다. 이 출력 벡터에는 개개인의 특징을 나타내는 외형 값과, 얼굴의 형태를 나타내는 표현 값으로 구성될 수 있다.
- [0030] 본 발명의 일 실시예에서는 3DMM을 이용하여 프레임 별 비디오 데이터로부터 3차원 포즈 파라미터(pose parameter), 기하 파라미터(geometry parameter) 및 표현 파라미터(expression parameter) 중 적어도 하나를 획득할 수 있다. 이를 통해, 서버(10)는 사람의 발화에 따른 입부분의 움직임뿐 아니라 얼굴의 전체적 상태를 포함하는 3차원 데이터를 획득할 수 있다.
- [0031] 다시 돌아가서, 도 3에 도시된 바와 같이, 서버(10)는 오디오 데이터를 Mel-frequency Cepstral Coefficients(MFCC) 모델에 입력하여 오디오 벡터(Fa)로 변환할 수 있다. 이때, 오디오 벡터를 추출하는 과정은 앞서 설명한 비디오 벡터를 추출하는 과정과 병렬적으로 수행될 수 있으며, 그 순서는 한정되지 않는다.
- [0032] MFCC란 음성 신호를 프레임 별로 나눈 후, 시간 단위에서 주파수 단위로 변환한 그래프이다. 본 발명의 일 실시예에서는 서버(10)가 MFCC 모델을 통해 영상의 오디오 데이터를 프레임 별로 나누어 고속 푸리에 변환(Fast

Fourier Transform, FFT)을 적용해 스펙트럼을 구하고, 구해진 스펙트럼에 멜 필터 뱅크(Mel Filter Bank)를 적용해 도 4와 같은 멜 스펙트럼(Mel Spectrum)을 구한 후, 멜 스펙트럼에 캡스트럴(Cepstral) 분석을 적용하여 MFCC를 추출한다.

- [0033] 이와 같은 MFCC 모델을 통해 음성 신호를 그대로 이용하는 것보다 훨씬 적은 양의 변수를 활용할 수 있고, 멜 스펙트럼을 사용함으로써 프레임 별로 시간에 대해 오디오 데이터가 정리되어 비디오 데이터와의 프레임 별 상관 관계를 용이하게 추출할 수 있다.
- [0034] 참고로, 서버(10)는 전자 장치(20)로부터 동영상에 대해 프레임 별로 비디오 벡터 및 오디오 벡터를 추출한 데이터를 수신할 수도 있으며, 이는 전자 장치(20)가 자체적으로 3DMM 모델을 이용한 비디오 벡터 추출 및 MFCC 모델을 이용한 오디오 벡터 추출 처리를 수행한 결과로서의 비디오 벡터 및 오디오 벡터를 서버(10)로 송신한 결과이다.
- [0035] 다시 도 2로 돌아가서, 서버(10)는 제1 기계 학습 모델을 통해 프레임 별 비디오 벡터 및 오디오 벡터 간의 상관 관계 값을 산출한다(S23).
- [0036] 구체적으로, 도 5에 도시된 바와 같이, 서버(10)는 제1 인코더(51)에 프레임 별 비디오 벡터 및 오디오 벡터를 결합하여 입력하고, 인코딩 결과로써 각 프레임에서 비디오 벡터와 오디오 벡터의 상관관계를 0~1사이의 실수로써 출력한다.
- [0037] 참고로, 제1 인코더(51)는 기계 학습 모델일 수 있으며, 기계 학습은 데이터를 분류하는 방식에 따라 대표적으로 의사결정나무 (Decision Tree)나 베이지안 망(Bayesian network), 서포트벡터머신(SVM: support vector machine), 및 인공 신경망(ANN: Artificial Neural Network) 등의 알고리즘으로 구분될 수 있다.
- [0038] 그 중 인공 신경망은 복수의 레이어(layer)를 포함하되, 레이어들 각각은 복수의 뉴런(neuron)을 포함하며. 또한 인공 신경망은 뉴런과 뉴런을 연결하는 시냅스를 포함할 수 있다. 일반적으로 인공 신경망은 세가지 인자, 즉 (1) 다른 레이어의 뉴런들 사이의 연결 패턴, (2) 연결의 가 중치를 갱신하는 학습 과정, (3) 이전 레이어로부터 수신되는 입력에 대한 가중합으로부터 출력값을 생성하는 활성화 함수에 의해 정의될 수 있다.
- [0039] 이러한 인공 신경망은 DNN(Deep Neural Network), RNN(Recurrent Neural Network), BRDNN(Bidirectional Recurrent Deep Neural Network), MLP(Multilayer Perceptron), CNN(Convolutional Neural Network)와 같은 방식의 네트 워크 모델들을 포함할 수 있으나, 이에 한정되지 않는다.
- [0040] 이때, 제1 인코더(51)의 구조는 인공 신경망의 FC 레이어(Fully Connected Layer)를 사용하며, 히든 레이어의 수가 점점 줄어들고 마지막에 소프트맥스(Softmax)를 사용해 0-1 사이의 실수값을 출력할 수 있다.
- [0041] 즉, 본 발명의 일 실시예에서 서버(10)는 비디오 벡터 및 오디오 벡터를 결합하여 인공 신경망의 FC 레이어(Fully Connected Layer)에 입력하고, FC 레이어를 통해 0과 1사이의 실수 값 중 각 프레임의 비디오 벡터 및 오디오 벡터 간의 상관 관계에 대응하는 실수 값을 출력할 수 있다. 이때, t번째 상관 관계 인코더(인코더t)의 출력 값(즉, 상관 관계 값)이 1에 가까울 수록 t번째 비디오 프레임과 오디오 프레임의 상관 관계가 높다는 것을 의미할 수 있다.
- [0042] 다시 도 2로 돌아가서, 서버(10)는 제2 기계 학습 모델을 통해 수신된 동영상에 페이크 비디오인지 여부를 판별한다(S24).
- [0043] 구체적으로 도 5에 도시된 바와 같이, 제 2 인코더(52)는 기계 학습 모델에 기반한 판별기 일 수 있으며, 인공 신경망을 사용할 수 있다. 서버(10)는 제1 인코더(51)로부터 출력된 결과 값을 결합하여 하나의 벡터로써 제2 인코더(52)에 입력할 수 있다.
- [0044] 이때, 제2 인코더(52)의 구조는 제1 인코더(51)와 유사하게 인공 신경망의 FC 레이어(Fully Connected Layer)를 사용하며, 최종 출력은 비디오 벡터와 오디오 벡터의 관계가 fake 인지 또는 real인지를 의미하는 0 또는 1 값을 출력한다.
- [0045] 즉, 본 발명의 일 실시예에서 서버(10)는 프레임 별 상관 관계 값을 결합한 벡터를 인공 신경망의 FC 레이어(Fully Connected Layer)에 입력하고, FC 레이어를 통해 동영상에 대한 비디오 벡터 및 오디오 벡터 간의 상관 관계가 페이크인지 여부를 나타내는 0 또는 1의 값을 출력할 수 있다.
- [0046] 한편, 서버(10)는 프레임 별 상관 관계 값과 기 설정된 기준 값을 비교한 결과에 따라 비디오 벡터 및 오디오 벡터 간의 관계가 페이크인지 여부를 판단하여, 동영상에 대해 프레임 별로 페이크 여부를 판단한 결과를 출력

할 수 있다. 즉, 제1 인코더(51)를 통해 출력된 비디오 벡터 및 오디오 벡터 간의 상관 관계 값은 0 내지 1 사이의 실수 값으로 표현되며, 제2 인코더(52)은 0 내지 1 사이의 실수 값을 갖는 상관 관계 값 중 소정의 값(즉, 기준 값) 이하의 값을 갖는 프레임에 대해서 페이크인 것으로 판별할 수 있다.

[0047] 또한, 서버(10)는 동영상에 포함된 복수의 프레임들의 비디오 벡터 및 오디오 벡터 간의 상관 관계 값(즉, 제1 인코더(51)의 출력 값)을 순차적으로 제2 인코더(52)에 입력하되, 기 설정된 프레임 개수 단위로 상기 상관 관계 값을 결합한 벡터를 제2 인코더(52)에 입력할 수 있다.

[0048] 또한, 서버(10)는 제2 인코더(52)를 통해 상기 상관 관계 값을 결합한 벡터가 최초로 페이크로 판별될 경우, 나머지 프레임들에 대한 제1 및 제2 인코더(51, 52)를 통한 검증 단계를 생략하거나 해당 검증 단계를 진행하는 것과는 별도로 해당 동영상을 페이크 비디오로써 우선 판별할 수 있다.

[0049] 다시 도 2로 돌아가서, 서버(10)는 수신된 영상에 대한 페이크 여부 판별 결과를 전자 장치(20)로 전송한다(S25).

[0050] 도 6은 본 발명의 일 실시예에 따라 도 2 내지 도 5의 동작을 서버(10)가 아닌 전자 장치(20)가 수행하는 페이크 영상 검증 시스템(2)을 도시한 일 예이다. 즉, 도 2 내지 도 5에서는 서버(10)가 도 2 내지 도 5의 방법을 수행하는 것으로 기술하였으나, 본 발명의 구현예에 따라, 도 6에 도시된 바와 같이 하나의 전자 장치(20)가 상기 도 2 내지 도 5의 방법을 수행할 수 있다.

[0051] 예를 들어, 하나의 전자 장치(20)는 복수의 비디오 프레임들을 포함하는 동영상으로부터 비디오 데이터 및 오디오 데이터를 추출하고, 프레임 별 비디오 및 오디오 데이터로부터 비디오 벡터 및 오디오 벡터를 추출하며, 프레임 별로 추출된 비디오 및 오디오 벡터 간의 상관 관계 값을 산출하고, 산출된 프레임 별 상관 관계 값에 기초하여 해당 동영상이 페이크 비디오인지 여부를 판별할 수 있다.

[0052] 또한, 구현예에 따라, 도 2 내지 도 5의 방법을 복수의 전자 장치들이 분산 수행하는 것도 가능하다는 것은 본 발명의 당업자에게 용이하게 이해될 수 있으므로 자세한 설명은 생략한다.

[0053] 도 7은 본 발명의 일 실시 예에 따른 서버(10)의 구성도이다.

[0054] 프로세서(710)는 하나 이상의 코어(core, 미도시) 및 그래픽 처리부(미도시) 및/또는 다른 구성 요소와 신호를 송수신하는 연결 통로(예를 들어, 버스(bus) 등)를 포함할 수 있다.

[0055] 일 실시예에 따른 프로세서(710)는 메모리(720)에 저장된 하나 이상의 인스트럭션을 실행함으로써, 도 1 내지 도 5와 관련하여 설명된 방법을 수행한다.

[0056] 한편, 프로세서(710)는 프로세서(710) 내부에서 처리되는 신호(또는, 데이터)를 일시적 및/또는 영구적으로 저장하는 램(RAM: Random Access Memory, 미도시) 및 롬(ROM: Read-Only Memory, 미도시)을 더 포함할 수 있다. 또한, 프로세서(710)는 그래픽 처리부, 램 및 롬 중 적어도 하나를 포함하는 시스템 온 칩(SoC: system on chip) 형태로 구현될 수 있다.

[0057] 메모리(720)에는 프로세서(710)의 처리 및 제어를 위한 프로그램들(하나 이상의 인스트럭션들)을 저장할 수 있다. 메모리(720)에 저장된 프로그램들은 기능에 따라 복수 개의 모듈들로 구분될 수 있다.

[0058] 또한, 서버(10)는 상기의 구성들 이외에, 각 기능에서 처리되는 데이터 등을 표시하고 사용자 입력(예컨대, 터치 등)을 수신하는 디스플레이부(미도시), 다른 전자 장치 등과 통신하기 위한 적어도 하나의 구성요소를 포함하는 통신부(미도시), 사용자의 음성을 입력받는 마이크(미도시) 등을 더 포함할 수 있다. 그러나, 본 발명의 서버(10)가 더 많은 구성요소들을 포함하여 구현될 수 있음은 본 발명의 기술분야의 당업자가 용이하게 이해할 수 있을 것이다.

[0059] 한편, 도 7의 구성요소들은 전자 장치(20)의 구성들에 대응될 수 있다. 즉, 전자 장치(20)는 하나 이상의 코어 및 그래픽 처리부를 포함하는 프로세서 및 메모리를 포함하며, 프로세서는 메모리에 저장된 하나 이상의 인스트럭션을 실행함으로써 도 1 내지 도 5와 관련하여 설명된 방법을 수행할 수 있다.

[0060] 한편, 본 발명의 실시예와 관련하여 설명된 방법 또는 알고리즘의 단계들은 하드웨어로 직접 구현되거나, 하드웨어에 의해 실행되는 소프트웨어 모듈로 구현되거나, 또는 이들의 결합에 의해 구현될 수 있다. 소프트웨어 모듈은 RAM(Random Access Memory), ROM(Read Only Memory), EPROM(Erasable Programmable ROM), EEPROM(Electrically Erasable Programmable ROM), 플래시 메모리(Flash Memory), 하드 디스크, 착탈형 디스크, CD-ROM, 또는 본 발명이 속하는 기술 분야에서 잘 알려진 임의의 형태의 컴퓨터 판독가능 기록매체에

상주할 수도 있다.

[0061] 본 발명의 구성 요소들은 하드웨어인 컴퓨터와 결합되어 실행되기 위해 프로그램(또는 애플리케이션)으로 구현되어 매체에 저장될 수 있다. 예를 들어, 본 발명의 일 실시예는 컴퓨터에서 독출가능한 기록매체에 저장된 컴퓨터프로그램으로서, 동영상으로부터 프레임 별 비디오 벡터 및 오디오 벡터를 추출하는 기능, 프레임 별로 추출된 비디오 벡터 및 오디오 벡터 간의 상관 관계 값을 산출하는 기능, 및 프레임 별로 산출된 상관 관계 값에 기초하여 동영상이 페이크 비디오인지 여부를 판별하는 기능을 수행할 수 있다.

[0062] 또한, 본 발명의 구성 요소들은 소프트웨어 프로그래밍 또는 소프트웨어 요소들로 실행될 수 있으며, 이와 유사하게, 실시 예는 데이터 구조, 프로세스들, 루틴들 또는 다른 프로그래밍 구성들의 조합으로 구현되는 다양한 알고리즘을 포함하여, C, C++, 자바(Java), 어셈블러(assembler), 파이썬(Python) 등과 같은 프로그래밍 또는 스크립팅 언어로 구현될 수 있다. 기능적인 측면들은 하나 이상의 프로세서들에서 실행되는 알고리즘으로 구현될 수 있다.

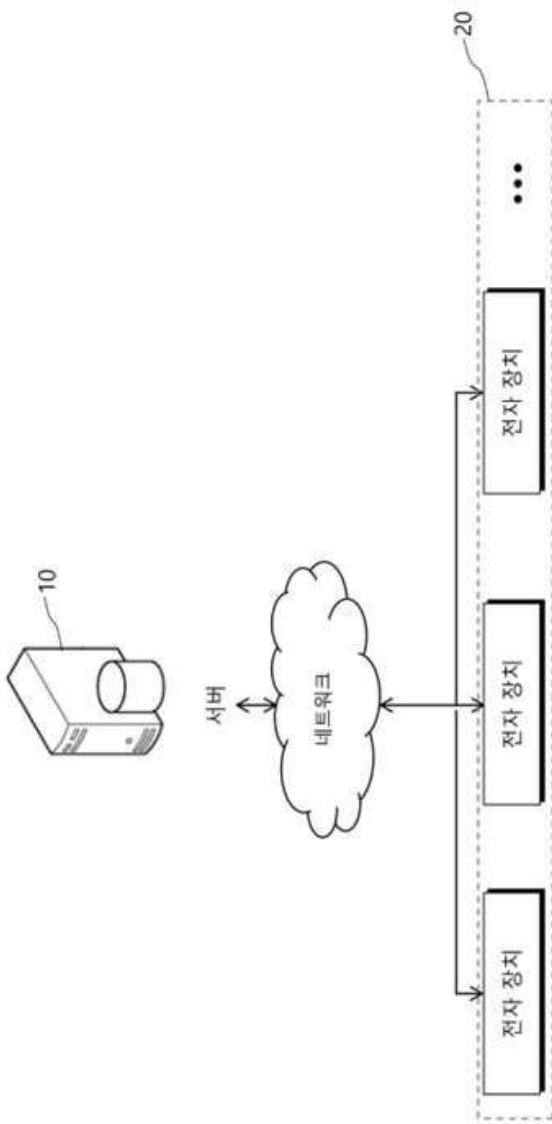
[0063] 이상의 설명은 본 발명의 기술 사상을 예시적으로 설명한 것에 불과한 것으로서, 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자라면 본 발명의 본질적인 특성에서 벗어나지 않는 범위에서 다양한 수정 및 변형이 가능할 것이다. 또한, 본 발명에 개시된 실시 예들은 본 발명의 기술 사상을 한정하기 위한 것이 아니라 설명하기 위한 것이고, 이러한 실시 예에 의하여 본 발명의 기술 사상의 범위가 한정되는 것은 아니다. 따라서, 본 발명의 보호 범위는 아래의 청구범위에 의하여 해석되어야 하며, 그와 동등한 범위 내에 있는 모든 기술 사상은 본 발명의 권리범위에 포함되는 것으로 해석되어야 할 것이다.

부호의 설명

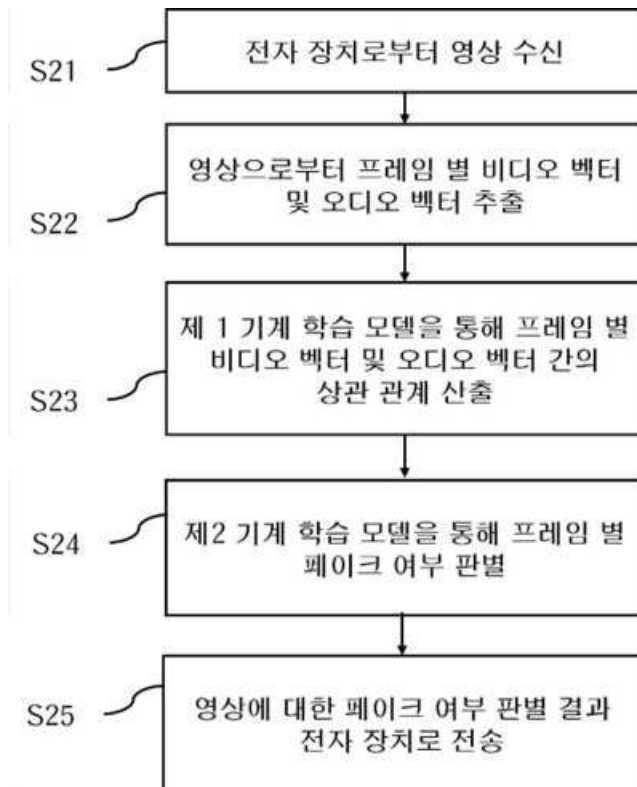
[0064] 1, 2: 페이크 영상 검증 시스템
10: 서버
20: 전자 장치
710: 프로세서
720: 메모리

도면

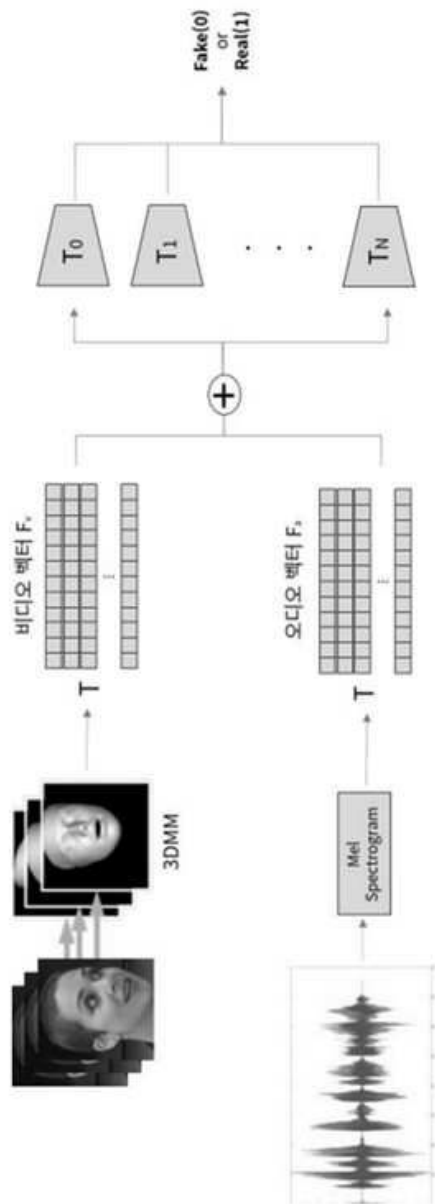
도면1



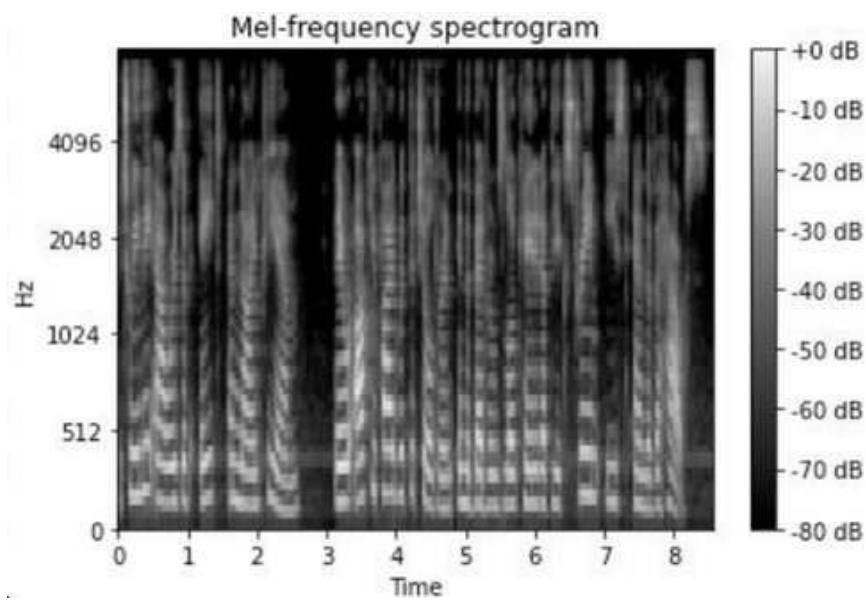
도면2



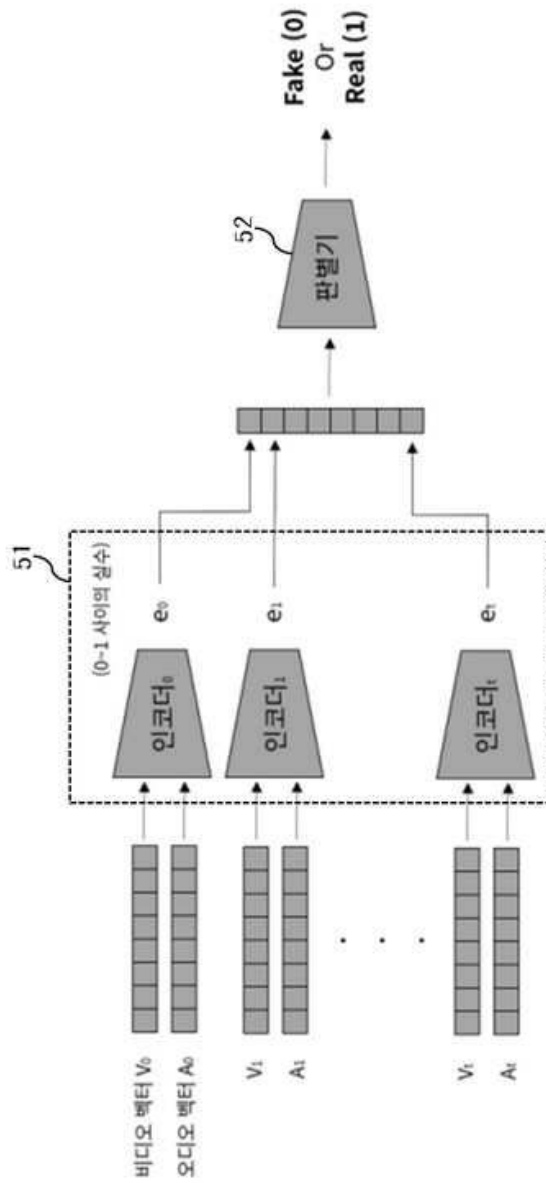
도면3



도면4

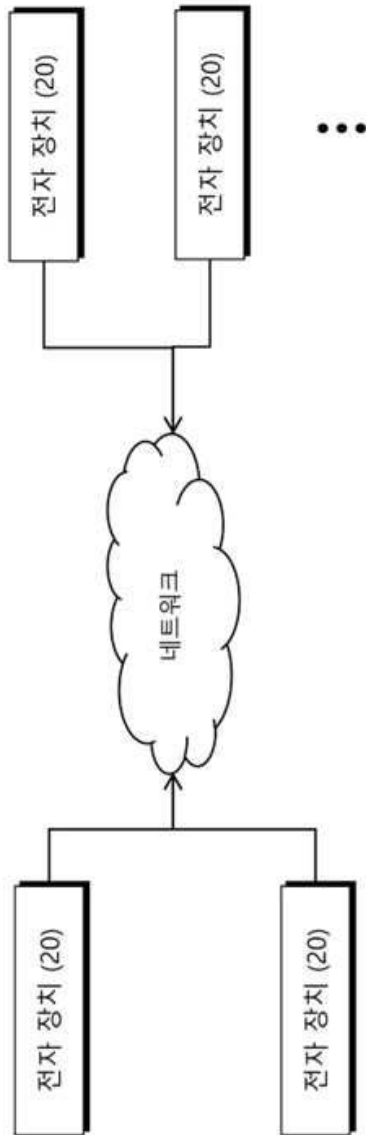


도면5



도면6

2



도면7

