



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2021년07월19일

(11) 등록번호 10-2279772

(24) 등록일자 2021년07월14일

(51) 국제특허분류(Int. Cl.)

G06T 19/00 (2011.01) G06T 3/00 (2019.01)

G06T 7/11 (2017.01)

(52) CPC특허분류

G06T 19/00 (2013.01)

G06T 3/00 (2019.01)

(21) 출원번호 10-2020-0010870

(22) 출원일자 2020년01월30일

심사청구일자 2020년01월30일

(56) 선행기술조사문헌

JP2000341602 A*

JP2019219914 A*

*는 심사관에 의하여 인용된 문헌

(73) 특허권자

연세대학교 산학협력단

서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)

(72) 발명자

변혜란

서울특별시 서대문구 연세로 50, 제4공학관 821호(신촌동, 연세대학교)

홍기범

서울특별시 서대문구 연세로 50, 제4공학관 821호(신촌동, 연세대학교)

김호성

서울특별시 서대문구 연세로 50, 제4공학관 821호(신촌동, 연세대학교)

(74) 대리인

특허법인우인

전체 청구항 수 : 총 14 항

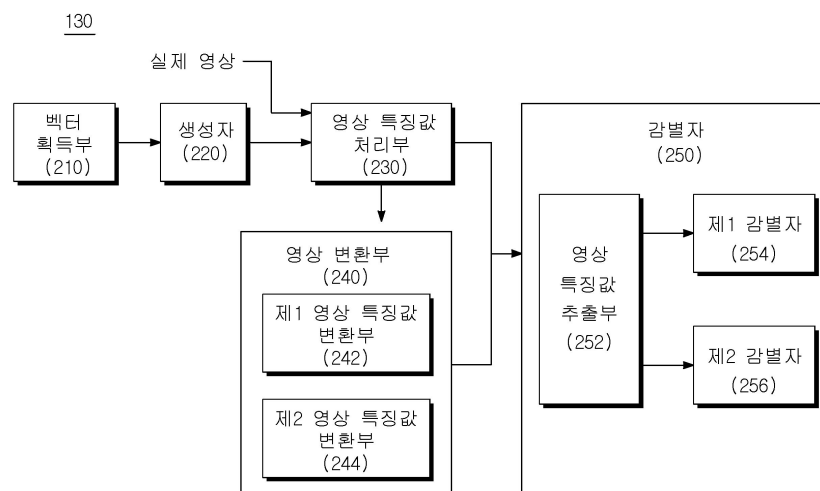
심사관 : 옥윤철

(54) 발명의 명칭 시간의 특성을 고려한 영상 생성 방법 및 그를 위한 장치

(57) 요약

시간의 특성을 고려한 영상 생성 방법 및 그를 위한 장치를 개시한다.

본 발명의 실시예에 따른 시간 특성을 고려한 영상 생성 학습 방법은, 소정의 입력값 기반의 잠재 벡터를 획득하는 벡터 획득 단계; 상기 잠재 벡터를 입력으로 가상 영상에 대한 가상 영상 데이터를 생성하는 생성 처리 단계; 실제 영상에 대한 실제 영상 데이터를 획득하는 영상 획득 단계; 상기 가상 영상 데이터 및 상기 실제 영상 데이터 각각의 시간 특성을 반전시켜 역방향 가상 영상 데이터 및 역방향 실제 영상 데이터를 생성하는 변환 단계; 및 상기 가상 영상 데이터, 상기 실제 영상 데이터, 상기 역방향 가상 영상 데이터 및 상기 역방향 실제 영상 데이터 중 적어도 두 개의 데이터를 비교하여 영상의 진위 여부 및 영상의 진행 방향 각각에 대한 분류를 처리하여 영상 생성이 수행되도록 하는 감별 처리 단계를 포함할 수 있다.

대표도

(52) CPC특허분류

G06T 7/11 (2017.01)

G06T 2207/20081 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	2019-11-0366
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	중견연구자지원사업
연구과제명	특성 정보 자동 생성을 통한 처음 보는 복합카테고리의 이미지와 비디오 생성 및 인
식을 위한 제로샷 학습 기술 연구(1/3)	
기 여 율	1/1
과제수행기관명	연세대학교
연구기간	2019.03.01 ~ 2020.02.29

명세서

청구범위

청구항 1

하나 이상의 프로세서 및 상기 프로세서에 의해 실행되는 하나 이상의 프로그램을 저장하는 메모리를 포함하는 컴퓨팅 디바이스에 의해 수행되는 영상 생성 학습 방법에 있어서, 상기 컴퓨팅 디바이스는,

소정의 입력값 기반의 잠재 벡터를 획득하는 벡터 획득 단계;

상기 잠재 벡터를 입력으로 가상 영상에 대한 가상 영상 데이터를 생성하는 생성 처리 단계;

실제 영상에 대한 실제 영상 데이터를 획득하는 영상 획득 단계;

상기 가상 영상 데이터 및 상기 실제 영상 데이터 각각의 시간 특성을 반전시켜 역방향 가상 영상 데이터 및 역방향 실제 영상 데이터를 생성하는 변환 단계; 및

상기 가상 영상 데이터, 상기 실제 영상 데이터, 상기 역방향 가상 영상 데이터 및 상기 역방향 실제 영상 데이터 중 적어도 두 개의 데이터를 비교하여 영상의 진위 여부 및 영상의 진행 방향 각각에 대한 분류를 처리하여 영상 생성이 수행되도록 하는 감별 처리 단계를 수행하되,

상기 생성 처리 단계는, 상기 잠재 벡터를 기반으로 생성된 시간이 연속되는 적어도 하나의 세그먼트 블록을 포함하는 상기 가상 영상 데이터를 생성하며, 상기 적어도 하나의 세그먼트 블록 각각은 RGB에 대한 특징값을 포함하는 것을 특징으로 하는 시간 특성을 고려한 영상 생성 학습 방법.

청구항 2

제1항에 있어서,

상기 벡터 획득 단계는,

기 설정된 개수의 정규분포 값으로 구성된 상기 잠재 벡터를 획득하고, 상기 잠재 벡터는 랜덤으로 선정된 상기 정규분포 값 또는 기 설정된 개수의 영상 프레임 각각에서 산출된 상기 정규분포 값으로 구성되는 것을 특징으로 하는 시간 특성을 고려한 영상 생성 학습 방법.

청구항 3

삭제

청구항 4

제1항에 있어서,

상기 영상 획득 단계는,

상기 실제 영상을 기반으로 생성된 시간이 연속되는 적어도 하나의 세그먼트 블록을 포함하는 상기 실제 영상 데이터를 입력 받으며, 상기 실제 영상 데이터에 포함된 상기 세그먼트 블록의 개수는 상기 가상 영상 데이터에 포함된 세그먼트 블록의 개수와 동일한 것을 특징으로 하는 시간 특성을 고려한 영상 생성 학습 방법.

청구항 5

제4항에 있어서,

상기 영상 획득 단계는,

상기 가상 영상과 무관한 상기 실제 영상에 대한 상기 실제 영상 데이터를 획득하며, 상기 실제 영상은 기 저장된 데이터베이스에서 추출되거나 네트워크 상에서 탐색된 영상의 전체 또는 일부에서 수집된 영상인 것을 특징으로 하는 시간 특성을 고려한 영상 생성 학습 방법.

청구항 6

제4항에 있어서,

상기 변환 단계는,

상기 실제 영상 데이터의 시간 특성을 기반으로 진행 순서를 반전시켜 상기 역방향 실제 영상 데이터를 생성하는 제1 변환 단계; 및

상기 가상 영상 데이터의 시간 특성을 기반으로 진행 순서를 반전시켜 상기 역방향 가상 영상 데이터를 생성하는 제2 변환 단계

를 포함하는 것을 특징으로 하는 시간 특성을 고려한 영상 생성 학습 방법.

청구항 7

제6항에 있어서,

상기 제1 변환 단계는,

상기 실제 영상 데이터에 포함된 상기 적어도 하나의 세그먼트 블록의 순서를 역순으로 변환하여 상기 역방향 실제 영상 데이터를 생성하는 것을 특징으로 하는 시간 특성을 고려한 영상 생성 학습 방법.

청구항 8

제6항에 있어서,

상기 제2 변환 단계는,

상기 가상 영상 데이터에 포함된 상기 적어도 하나의 세그먼트 블록의 순서를 역순으로 변환하여 상기 역방향 가상 영상 데이터를 생성하는 것을 특징으로 하는 시간 특성을 고려한 영상 생성 학습 방법.

청구항 9

제4항에 있어서,

상기 감별 처리 단계는,

상기 가상 영상 데이터, 상기 실제 영상 데이터, 상기 역방향 가상 영상 데이터 및 상기 역방향 실제 영상 데이터 각각에 대한 특징값을 추출하는 특징값 추출 단계;

상기 가상 영상 데이터 및 상기 실제 영상 데이터에 대한 특징값을 이용하여 상기 가상 영상의 진위 여부에 대한 분류를 처리하는 제1 감별 단계; 및

상기 가상 영상 데이터 및 상기 역방향 가상 영상 데이터에 대한 특징값을 이용하여 영상의 진행 방향에 대한 분류를 처리하는 제2 감별 단계

를 포함하는 것을 특징으로 하는 시간 특성을 고려한 영상 생성 학습 방법.

청구항 10

제9항에 있어서,

상기 제1 감별 단계 및 상기 제2 감별 단계는,

서로 공유된 적어도 하나의 컨볼루션 레이어를 통과한 상기 특징값을 기반으로 영상의 진위 여부 및 영상의 진행 방향 각각에 대한 분류를 처리하는 것을 특징으로 하는 시간 특성을 고려한 영상 생성 학습 방법.

청구항 11

제9항에 있어서,

상기 제1 감별 단계는,

상기 가상 영상 데이터 및 상기 실제 영상 데이터의 특징값을 비교하여 상기 가상 영상의 진위 여부를 학습한 제1 학습 결과를 출력하되,

상기 제1 감별 단계는, 상기 제1 학습 결과에 근거하여 상기 생성 처리 단계로 피드백 정보를 전달하며, 상기 생성 처리 단계에서 상기 가상 영상이 참 신호에 해당하는 가상 영상이 생성될 때까지 반복하여 상기 가상 영상의 진위 여부를 학습하는 것을 특징으로 하는 시간 특성을 고려한 영상 생성 학습 방법.

청구항 12

제9항에 있어서,

상기 제1 감별 단계는,

상기 가상 영상 데이터 및 상기 실제 영상 데이터의 특징값을 비교하여 상기 가상 영상의 진위 여부를 학습한 제1 학습 결과를 출력하고,

상기 제2 감별 단계는, 상기 가상 영상 데이터 및 상기 역방향 가상 영상 데이터의 특징값을 비교하여 상기 가상 영상의 진행 방향을 학습한 제2 학습 결과를 출력하되,

상기 제1 학습 결과 및 상기 제2 학습 결과를 기반으로 피드백 정보를 산출하고, 상기 피드백 정보를 이용하여 상기 감별 처리 단계에서 상기 가상 영상 데이터 및 상기 실제 영상 데이터를 비교하여 상기 가상 영상이 참 신호에 해당할 때까지 반복하여 상기 영상의 진위 여부를 학습하는 것을 특징으로 하는 시간 특성을 고려한 영상 생성 학습 방법.

청구항 13

시간 특성을 고려하여 영상을 생성하는 장치로서,

하나 이상의 프로세서; 및

상기 프로세서에 의해 실행되는 하나 이상의 프로그램을 저장하는 메모리를 포함하며, 상기 프로그램들은 하나 이상의 프로세서에 의해 실행될 때, 상기 하나 이상의 프로세서들에서,

소정의 입력값 기반의 잠재 벡터를 획득하는 벡터 획득 단계;

상기 잠재 벡터를 입력으로 가상 영상에 대한 가상 영상 데이터를 생성하는 생성 처리 단계;

실제 영상에 대한 실제 영상 데이터를 획득하는 영상 획득 단계;

상기 가상 영상 데이터 및 상기 실제 영상 데이터 각각의 시간 특성을 반전시켜 역방향 가상 영상 데이터 및 역방향 실제 영상 데이터를 생성하는 변환 단계; 및

상기 가상 영상 데이터, 상기 실제 영상 데이터, 상기 역방향 가상 영상 데이터 및 상기 역방향 실제 영상 데이터 중 적어도 두 개의 데이터를 비교하여 영상의 진위 여부 및 영상의 진행 방향 각각에 대한 분류를 처리하여 영상 생성이 수행되도록 하는 감별 처리 단계를 포함하는 동작들을 수행하게 하되,

상기 생성 처리 단계는, 상기 잠재 벡터를 기반으로 생성된 시간이 연속되는 적어도 하나의 세그먼트 블록을 포함하는 상기 가상 영상 데이터를 생성하며, 상기 적어도 하나의 세그먼트 블록 각각은 RGB에 대한 특징값을 포함하는 것을 특징으로 하는 영상 생성 장치.

청구항 14

삭제

청구항 15

제13항에 있어서,

상기 영상 획득 단계는,

상기 실제 영상을 기반으로 생성된 시간이 연속되는 적어도 하나의 세그먼트 블록을 포함하는 상기 실제 영상 데이터를 입력 받으며, 상기 실제 영상 데이터에 포함된 상기 세그먼트 블록의 개수는 상기 가상 영상 데이터에 포함된 세그먼트 블록의 개수와 동일한 것을 특징으로 하는 영상 생성 장치.

청구항 16

제15항에 있어서,

상기 변환 단계는,

상기 실제 영상 데이터의 시간 특성을 기반으로 진행 순서를 반전시켜 상기 역방향 실제 영상 데이터를 생성하는 제1 변환 단계; 및

상기 가상 영상 데이터의 시간 특성을 기반으로 진행 순서를 반전시켜 상기 역방향 가상 영상 데이터를 생성하는 제2 변환 단계

를 포함하는 것을 특징으로 하는 영상 생성 장치.

청구항 17

삭제

발명의 설명

기술 분야

[0001] 본 발명은 시간의 특성을 고려하여 영상을 생성하기 위한 방법 및 그를 위한 장치에 관한 것이다.

배경 기술

[0002] 이 부분에 기술된 내용은 단순히 본 발명의 실시예에 대한 배경 정보를 제공할 뿐 종래기술을 구성하는 것은 아니다.

[0003] 기계학습(machine learning)은 지도학습(supervised learning), 비지도 학습(unsupervised learning) 및 강화 학습(reinforcement learning) 등으로 나누어 진다. 비지도 학습은 지도학습과는 달리 목표값이 주어지지 않는 자료를 학습하여 차원 축소, 군집화, 특징추출 등의 작업을 주로 하는데 오토인코더(autoencoder), GAN(generative adversarial networks), RBM(restricted Boltzmann machine) 등이 있다.

[0004] 딥러닝 분야에서는 지도학습이 많은 부분을 차지했지만, 목표값(target)을 구해야 하는 이유로 활동에 한계가 있었다. 반면, 비지도학습은 Goodfellow(2014)의 GAN을 중심으로 넓은 활용도를 보이며 현재까지 영상(비디오), 이미지와 음성 분야를 비롯한 다양한 분야에서 접목되고 있다.

[0005] GAN은 생성자(generator)와 감별자(discriminator)를 대립하여 학습하는 신경망이다. 생성자는 실제 자료와 같은 자료를 만들기 위해 노력하고, 감별자는 실제 자료와 생성자료를 구분하기 위해 노력하여 최종적으로 생성자가 실제 자료와 같은 자료를 만드는 것이 GAN의 목표다. 초기 GAN은 학습의 불안정과 모드 붕괴(mode-collapse) 등의 문제점을 수반하였다. 이를 개선하는 많은 파생 GAN 모형들이 개발되었으며, 인셉션 모델(Inception model)을 이용한 인셉션 스코어(IS: Inception score)와 프레첷 인셉션 거리(FID: Frechet Inception distance) 등이 개발되어 GAN의 성능을 평가할 수 있게 되었다.

[0006] 기존의 영상(비디오) 생성에 대한 연구들은 영상의 시간적 특성에 대해 크게 고려하지 않은 채 GAN의 학습 방식만을 토대로 영상 생성 연구를 진행하였다. 영상(비디오)은 시간을 구별하는 차원을 가지고 있기 때문에, 영상에 대한 GAN 학습은 이미지에 대한 GAN 학습보다 정교하다. 시간을 고려한 전용 아키텍처를 설계하는 방법이 연구되고 있으나, 생성된 영상은 여전히 실제 영상과 구별을 할 수 없다.

[0007] 바닐라 GAN에서 StyleGAN에 이르기까지 이미지 생성에 대한 학습이 엄격하게 연구되었지만, 추가적인 특성(시간적 특성)으로 인해 영상 제작에 어려움이 있다.

[0008] 영상 생성을 위해 개발된 VGAN은 다른 차원과 마찬가지로 시간을 고려한 3D 컨볼루션이 있는 생성자와 감별자로 구성된다. 그 후, TGAN과 MoCoGAN은 생성 과정에서 시간적 일관성을 모델링하기 위한 주요 성분으로 일시적 잠재 변수를 도입했다. 그러나, 이러한 GAN 방식에서 생성된 영상은 생성된 이미지에 비해 정교하지 못하다.

[0009] 이에, 시간의 특성(AoT: Arrow of Time)을 고려하여 영상을 생성하는 모델을 통해 정교하게 영상을 생성하기 위한 방법이 필요하다.

발명의 내용

해결하려는 과제

- [0010] 본 발명은 실제 영상과 생성된 가상 영상뿐만 아니라, 실제 영상 및 가상 영상의 시간 특성을 반전시켜 역방향으로 변환된 영상을 분류를 처리하는 학습을 통해 시간의 특성을 고려한 영상 생성 방법 및 그를 위한 장치를 제공하는 데 주된 목적이 있다.

과제의 해결 수단

- [0011] 본 발명의 일 측면에 의하면, 상기 목적을 달성하기 위한 하나 이상의 프로세서 및 상기 프로세서에 의해 실행되는 하나 이상의 프로그램을 저장하는 메모리를 포함하는 컴퓨팅 디바이스에 의해 수행되는 시간 특성을 고려한 영상 생성 학습 방법은, 소정의 입력값 기반의 잠재 벡터를 획득하는 벡터 획득 단계; 상기 잠재 벡터를 입력으로 가상 영상에 대한 가상 영상 데이터를 생성하는 생성 처리 단계; 실제 영상에 대한 실제 영상 데이터를 획득하는 영상 획득 단계; 상기 가상 영상 데이터 및 상기 실제 영상 데이터 각각의 시간 특성을 반전시켜 역방향 가상 영상 데이터 및 역방향 실제 영상 데이터를 생성하는 변환 단계; 및 상기 가상 영상 데이터, 상기 실제 영상 데이터, 상기 역방향 가상 영상 데이터 및 상기 역방향 실제 영상 데이터 중 적어도 두 개의 데이터를 비교하여 영상의 진위 여부 및 영상의 진행 방향 각각에 대한 분류를 처리하여 영상 생성이 수행되도록 하는 감별 처리 단계를 수행할 수 있다.
- [0012] 또한, 본 발명의 다른 측면에 의하면, 상기 목적을 달성하기 위한 영상 생성 장치는, 시간 특성을 고려하여 영상을 생성하는 장치로서, 하나 이상의 프로세서; 및 상기 프로세서에 의해 실행되는 하나 이상의 프로그램을 저장하는 메모리를 포함하며, 상기 프로그램들은 하나 이상의 프로세서에 의해 실행될 때, 상기 하나 이상의 프로세서들에서, 소정의 입력값 기반의 잠재 벡터를 획득하는 벡터 획득 단계; 상기 잠재 벡터를 입력으로 가상 영상에 대한 가상 영상 데이터를 생성하는 생성 처리 단계; 실제 영상에 대한 실제 영상 데이터를 획득하는 영상 획득 단계; 상기 가상 영상 데이터 및 상기 실제 영상 데이터 각각의 시간 특성을 반전시켜 역방향 가상 영상 데이터 및 역방향 실제 영상 데이터를 생성하는 변환 단계; 및 상기 가상 영상 데이터, 상기 실제 영상 데이터, 상기 역방향 가상 영상 데이터 및 상기 역방향 실제 영상 데이터 중 적어도 두 개의 데이터를 비교하여 영상의 진위 여부 및 영상의 진행 방향 각각에 대한 분류를 처리하여 영상 생성이 수행되도록 하는 감별 처리 단계를 포함하는 동작들을 수행할 수 있다.
- [0013] 또한, 본 발명의 다른 측면에 의하면, 상기 목적을 달성하기 위한 하나 이상의 프로세서 및 상기 프로세서에 의해 실행되는 하나 이상의 프로그램을 저장하는 메모리를 포함하는 컴퓨팅 디바이스에 의해 수행되는 영상 생성 방법은, 입력 벡터를 입력 받고, 상기 입력 벡터에 대한 특징값을 추출하며, 가상 영상 데이터 및 실제 영상 데이터의 특징값을 비교하여 가상 영상의 진위 여부를 학습한 제1 학습 결과와 가상 영상 데이터 및 역방향 가상 영상 데이터의 특징값을 비교하여 가상 영상의 진행 방향을 학습한 제2 학습 결과를 적용하여 신규 영상을 생성하고, 생성된 신규 영상을 출력할 수 있다.

발명의 효과

- [0014] 이상에서 설명한 바와 같이, 본 발명은 시간적 특성을 고려하여 사실적인 영상 생성이 가능한 효과가 있다.
- [0015] 또한, 본 발명은 Self-supervision을 활용하여 다양한 데이터 셋을 자동으로 구축할 수 있는 효과가 있다.
- [0016] 또한, 본 발명은 데이터셋을 직접 수작업으로 수정할 필요 없이 자동으로 정답지가 생성됨에 따라 사람의 노동력(human labor)도 줄이면서 학습에 더 효율적인 성능을 이끌어 낼 수 있는 효과가 있다.

도면의 간단한 설명

- [0017] 도 1은 본 발명의 실시예에 따른 영상 생성 장치를 개략적으로 나타낸 블록 구성도이다.
- 도 2는 본 발명의 실시예에 따른 프로세서의 학습을 위한 동작 구성을 개략적으로 나타낸 블록 구성도이다.
- 도 3은 본 발명의 실시예에 따른 영상 생성을 위한 학습 방법을 설명하기 위한 순서도이다.
- 도 4는 본 발명의 실시예에 따른 프로세서의 영상 생성을 위한 동작 구성을 개략적으로 나타낸 블록 구성도이다.
- 도 5는 본 발명의 실시예에 따른 영상 생성 방법을 설명하기 위한 순서도이다.

도 6은 본 발명의 제1 실시예에 따른 영상 생성 장치의 학습 동작을 설명하기 위한 예시도이다.

도 7은 본 발명의 제2 실시예에 따른 영상 생성 장치의 학습 동작을 설명하기 위한 예시도이다.

도 8a 내지 도 8c는 본 발명의 실시예에 따른 영상 생성 장치의 학습 결과 및 적용 결과를 나타낸 도면이다.

발명을 실시하기 위한 구체적인 내용

- [0018] 이하, 본 발명의 바람직한 실시예를 첨부된 도면들을 참조하여 상세히 설명한다. 본 발명을 설명함에 있어, 관련된 공지 구성 또는 기능에 대한 구체적인 설명이 본 발명의 요지를 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명은 생략한다. 또한, 이하에서 본 발명의 바람직한 실시예를 설명할 것이나, 본 발명의 기술적 사상은 이에 한정하거나 제한되지 않고 당업자에 의해 변형되어 다양하게 실시될 수 있음은 물론이다. 이하에서는 도면들을 참조하여 본 발명에서 제안하는 시간의 특성을 고려한 영상 생성 방법 및 그를 위한 장치에 대해 자세하게 설명하기로 한다.
- [0019] 도 1은 본 발명의 실시예에 따른 영상 생성 장치를 개략적으로 나타낸 블록 구성도이다.
- [0020] 본 실시예에 따른 영상 생성 장치(100)는 입력부(110), 출력부(120), 프로세서(130), 메모리(140) 및 데이터 베이스(150)를 포함한다. 도 1의 영상 생성 장치(100)는 일 실시예에 따른 것으로서, 도 1에 도시된 모든 블록이 필수 구성요소는 아니며, 다른 실시예에서 영상 생성 장치(100)에 포함된 일부 블록이 추가, 변경 또는 삭제될 수 있다. 한편, 영상 생성 장치(100)는 컴퓨팅 디바이스로 구현될 수 있고, 영상 생성 장치(100)에 포함된 각 구성요소들은 각각 별도의 소프트웨어 장치로 구현되거나, 소프트웨어가 결합된 별도의 하드웨어 장치로 구현될 수 있다.
- [0021] 영상 생성 장치(100)는 잠재 벡터를 입력 받고, 잠재 벡터를 입력으로 생성자를 통해 가상 영상에 대한 가상 영상 데이터를 생성하면서 실제 영상에 대한 실제 영상 데이터를 획득하며, 가상 영상 데이터 및 실제 영상 데이터 각각의 시간 특성을 반전시켜 역방향 가상 영상 데이터 및 역방향 실제 영상 데이터를 생성하고, 생성자와 연동하는 감별자를 통해 가상 영상 데이터, 실제 영상 데이터, 역방향 가상 영상 데이터 및 역방향 실제 영상 데이터 중 적어도 두 개의 데이터를 비교하여 영상의 진위 여부 및 영상의 진행 방향 각각에 대한 분류를 처리하여 영상 생성이 수행되도록 하는 동작을 수행한다.
- [0022] 입력부(110)는 영상 생성 장치(100)에서의 영상 생성 동작을 수행하기 위한 신호 또는 데이터를 입력하거나 획득하는 수단을 의미한다. 입력부(110)는 프로세서(130)와 연동하여 다양한 형태의 신호 또는 데이터를 입력하거나, 외부 장치와의 연동을 통해 신호 또는 데이터를 획득하여 프로세서(130)로 전달할 수도 있다. 여기서, 입력부(110)는 실제 영상, 잠재 벡터, 랜덤 변수, 소정의 입력값 등을 입력하기 위한 모듈로 구현될 수 있으나 반드시 이에 한정되는 것은 아니다.
- [0023] 출력부(120)는 프로세서(130)와 연동하여 가상 영상 데이터 기반의 영상의 진위 여부 및 영상의 진행 방향 등의 학습 결과, 영상 생성 결과 등 다양한 정보를 출력할 수 있다. 출력부(120)는 영상 생성 장치(100)에 구비된 디스플레이(미도시)를 통해 다양한 정보를 출력할 수 있으나 반드시 이에 한정되는 것은 아니며, 다양한 형태의 방식으로 출력을 수행할 수 있다.
- [0024] 프로세서(130)는 메모리(140)에 포함된 적어도 하나의 명령어 또는 프로그램을 실행시키는 기능을 수행한다.
- [0025] 본 실시예에 따른 프로세서(130)는 입력부(110) 또는 데이터 베이스(150)로부터 획득한 잠재 벡터 또는 실제 영상 데이터를 기반으로 기계학습을 수행하고, 기계학습 결과를 기반으로 기 학습되지 않은 가상 영상을 생성하는 동작을 수행한다.
- [0026] 프로세서(130)는 획득된 잠재 벡터를 입력으로 가상 영상에 대한 가상 영상 데이터를 생성하고, 실제 영상에 대한 실제 영상 데이터를 획득한다.
- [0027] 또한, 프로세서(130)는 가상 영상 데이터 및 실제 영상 데이터 각각의 시간 특성을 반전시켜 역방향 가상 영상 데이터 및 역방향 실제 영상 데이터를 생성한다.
- [0028] 또한, 프로세서(130)는 가상 영상 데이터, 실제 영상 데이터, 역방향 가상 영상 데이터 및 역방향 실제 영상 데이터 중 적어도 두 개의 데이터를 비교하여 영상의 진위 여부 및 영상의 진행 방향 각각에 대한 분류를 처리하여 영상 생성이 수행되도록 한다. 본 실시예에 따른 프로세서(130)의 자세한 동작은 도 2 내지 5에서 설명하도록 한다.

- [0029] 메모리(140)는 프로세서(130)에 의해 실행 가능한 적어도 하나의 명령어 또는 프로그램을 포함한다. 메모리(140)는 가상 영상 데이터를 생성하는 동작, 영상 데이터의 진행 방향을 변환하는 동작, 가상 영상에 대한 진위 여부의 분류를 처리하는 동작, 가상 영상에 대한 진행 방향의 분류를 처리하는 동작, 영상 생성을 수행하는 동작 등을 위한 명령어 또는 프로그램을 포함할 수 있다.
- [0030] 데이터베이스(150)는 데이터베이스 관리 프로그램(DBMS)을 이용하여 컴퓨터 시스템의 저장공간(하드디스크 또는 메모리)에 구현된 일반적인 데이터구조를 의미하는 것으로, 데이터의 검색(추출), 삭제, 편집, 추가 등을 자유롭게 행할 수 있는 데이터 저장형태를 뜻하는 것으로, 오라클(Oracle), 인포믹스(Infomix), 사이베이스(Sybase), DB2와 같은 관계형 데이터베이스 관리 시스템(RDBMS)이나, 겔스톤(Gemston), 오리온(Orion), 02 등과 같은 객체 지향 데이터베이스 관리 시스템(OODBMS) 및 엑셀론(Excelon), 타미노(Tamino), 세카이주(Sekaiju) 등의 XML 전용 데이터베이스(XML Native Database)를 이용하여 본 발명의 일 실시예의 목적에 맞게 구현될 수 있고, 자신의 기능을 달성하기 위하여 적당한 필드(Field) 또는 엘리먼트들을 가지고 있다.
- [0031] 본 실시예에 따른 데이터베이스(150)는 영상 생성 또는 영상 생성을 위한 학습과 관련된 데이터를 저장하고, 기 저장된 영상 생성 또는 영상 생성을 위한 학습과 관련된 데이터를 제공할 수 있다.
- [0032] 데이터베이스(150)에 저장된 데이터는 잠재 벡터, 영상 데이터(예: 실제 영상 데이터, 가상 영상 데이터, 역방향 실제 영상 데이터, 역방향 가상 영상 데이터 등), 학습 결과(예: 제1 학습 결과, 제2 학습 결과, 영상 생성 학습 결과 등), 영상 생성 결과 등에 대한 데이터일 수 있다. 데이터베이스(140)는 영상 생성 장치(100) 내에 구현되는 것으로 기재하고 있으나 반드시 이에 한정되는 것은 아니며, 별도의 데이터 저장장치로 구현될 수도 있다.
- [0033] 도 2는 본 발명의 실시예에 따른 프로세서의 학습을 위한 동작 구성을 개략적으로 나타낸 블록 구성도이다.
- [0034] 본 실시예에 따른 영상 생성 장치(100)에 포함된 프로세서(130)는 기계 학습을 기반으로 영상을 생성하는 동작을 수행한다. 여기서, 기계 학습은 생성적 적대 신경망(GAN: Generative Adversarial Network)을 이용한 학습인 것이 바람직하나 반드시 이에 한정되는 것은 아니다.
- [0035] 영상 생성 장치(100)에 포함된 프로세서(130)는 획득된 잠재 벡터를 입력으로 가상 영상에 대한 가상 영상 데이터를 생성하고, 실제 영상에 대한 실제 영상 데이터를 획득하며, 가상 영상 데이터 및 실제 영상 데이터 각각의 시간 특성을 반전시켜 역방향 가상 영상 데이터 및 역방향 실제 영상 데이터를 생성하고, 가상 영상 데이터, 실제 영상 데이터, 역방향 가상 영상 데이터 및 역방향 실제 영상 데이터 중 적어도 두 개의 데이터를 비교하여 영상의 진위 여부 및 영상의 진행 방향 각각에 대한 분류를 처리하여 영상 생성이 수행되도록 하는 동작이 수행되도록 하며, 영상 생성을 수행하는 모든 기기에 탑재되거나, 영상 생성을 수행하는 소프트웨어와 연동할 수 있다.
- [0036] 본 실시예에 따른 프로세서(130)는 벡터 획득부(210), 생성자(220), 영상 특징값 처리부(230), 영상 변환부(240), 감별자(250)를 포함할 수 있다. 도 2의 프로세서(130)는 일 실시예에 따른 것으로서, 도 2에 도시된 모든 블록이 필수 구성요소는 아니며, 다른 실시예에서 프로세서(130)에 포함된 일부 블록이 추가, 변경 또는 삭제될 수 있다. 한편, 프로세서(130)에 포함된 각 구성요소들은 각각 별도의 소프트웨어 장치로 구현되거나, 소프트웨어가 결합된 별도의 하드웨어 장치로 구현될 수 있다.
- [0037] 벡터 획득부(210)는 소정의 입력값 기반의 잠재 벡터를 획득한다.
- [0038] 벡터 획득부(210)는 기 설정된 개수의 정규분포 값으로 구성된 잠재 벡터를 획득한다. 여기서, 잠재 벡터는 랜덤으로 선정된 정규분포 값 또는 기 설정된 개수의 영상 프레임 각각에서 산출된 정규분포 값으로 구성될 수 있다.
- [0039] 생성자(220)는 잠재 벡터를 입력으로 가상 영상에 대한 가상 영상 데이터를 생성하는 동작을 수행한다.
- [0040] 생성자(220)는 잠재 벡터를 기반으로 생성된 시간이 연속되는 적어도 하나의 세그먼트 블록을 포함하는 가상 영상 데이터를 생성한다. 여기서, 가상 영상 데이터에 포함된 상기 적어도 하나의 세그먼트 블록 각각은 RGB에 대한 특징값을 포함할 수 있다. 여기서, 생성자(220)는 컨볼루션 뉴럴 네트워크(CNN: Convolutional Neural Networks) 학습을 통해 가상 영상 데이터를 생성하는 것이 바람직하나 반드시 이에 한정되는 것은 아니다.
- [0041] 영상 특징값 처리부(230)는 생성자(220)로부터 가상 영상 데이터를 획득하고, 외부 장치로부터 실제 영상에 대한 실제 영상 데이터를 획득한다.

- [0042] 영상 특징값 처리부(230)는 실제 영상을 기반으로 생성된 시간이 연속되는 적어도 하나의 세그먼트 블록을 포함하는 실제 영상 데이터를 입력 받을 수 있다. 여기서, 실제 영상 데이터에 포함된 상기 세그먼트 블록의 개수는 상기 가상 영상 데이터에 포함된 세그먼트 블록의 개수와 동일한 것이 바람직하다.
- [0043] 영상 특징값 처리부(230)는 가상 영상과 무관한 실제 영상에 대한 실제 영상 데이터를 획득할 수 있다. 여기서, 실제 영상은 기 저장된 데이터베이스에서 추출되거나 네트워크 상에서 탐색된 영상의 전체 또는 일부에서 수집된 영상일 수 있다.
- [0044] 영상 특징값 처리부(230)는 획득된 가상 영상 데이터 및 실제 영상 데이터를 감별자(250)로 전달하는 동작을 수행한다. 한편, 생성자(220)에서 가상 영상 데이터 및 실제 영상 데이터를 감별자(250)로 직접 전달하는 경우, 영상 특징값 처리부(230)는 생략되거나, 생성자(220)에 포함된 형태로 구현될 수 있다.
- [0045] 영상 변환부(240)는 가상 영상 데이터 및 실제 영상 데이터를 입력 받고, 가상 영상 데이터 및 실제 영상 데이터 각각의 시간 특성을 반전시켜 역방향 가상 영상 데이터 및 역방향 실제 영상 데이터를 생성하는 동작을 수행한다.
- [0046] 본 실시예에 따른 영상 변환부(240)는 제1 영상 특징값 변환부(242) 및 제2 영상 특징값 변환부(244)를 포함한다.
- [0047] 제1 영상 특징값 변환부(242)는 실제 영상 데이터의 시간 특성을 기반으로 진행 순서를 반전시켜 상기 역방향 실제 영상 데이터를 생성하는 동작을 수행한다. 구체적으로, 제1 영상 특징값 변환부(242)는 실제 영상 데이터에 포함된 적어도 하나의 세그먼트 블록의 순서를 역순으로 변환하여 역방향 실제 영상 데이터를 생성할 수 있다.
- [0048] 제2 영상 특징값 변환부(244)는 가상 영상 데이터의 시간 특성을 기반으로 진행 순서를 반전시켜 역방향 가상 영상 데이터를 생성하는 동작을 수행한다. 구체적으로, 제2 영상 특징값 변환부(244)는 가상 영상 데이터에 포함된 적어도 하나의 세그먼트 블록의 순서를 역순으로 변환하여 역방향 가상 영상 데이터를 생성할 수 있다.
- [0049] 감별자(250)는 영상 특징값 처리부(230) 및 영상 변환부(240)과 연동하여 영상 생성을 위한 감별 처리를 수행한다.
- [0050] 감별자(250)는 가상 영상 데이터, 실제 영상 데이터, 역방향 가상 영상 데이터 및 역방향 실제 영상 데이터 중 적어도 두 개의 데이터를 비교하여 영상의 진위 여부 및 영상의 진행 방향 각각에 대한 분류를 처리하여 영상 생성이 수행되도록 한다.
- [0051] 본 실시예에 따른 감별자(250)는 영상 특징값 추출부(252), 제1 감별자(254) 및 제2 감별자(256)를 포함한다.
- [0052] 영상 특징값 추출부(252)는 가상 영상 데이터, 실제 영상 데이터, 역방향 가상 영상 데이터 및 역방향 실제 영상 데이터 각각에 대한 특징값을 추출할 수 있다. 영상 특징값 추출부(252)는 제1 감별자(254) 및 제2 감별자(256)와 서로 공유된 적어도 하나의 컨볼루션 레이어로 구성될 수 있다.
- [0053] 제1 감별자(254) 및 제2 감별자(256)는 서로 공유된 적어도 하나의 컨볼루션 레이어를 통과한 특징값을 기반으로 영상의 진위 여부 및 영상의 진행 방향 각각에 대한 분류를 처리한다.
- [0054] 제1 감별자(254)는 가상 영상 데이터 및 실제 영상 데이터에 대한 특징값을 이용하여 상기 가상 영상의 진위 여부에 대한 분류를 처리한다. 제1 감별자(254)는 가상 영상 데이터 및 실제 영상 데이터의 특징값을 비교하여 가상 영상의 진위 여부를 학습하고, 학습한 제1 학습 결과를 출력한다.
- [0055] 제1 감별자(254)는 제1 학습 결과에 근거하여 생성자(220)로 피드백 정보를 전달하며, 생성자(220)에서 가상 영상이 참 신호로 분류되는 가상 영상이 생성될 때까지 반복하여 가상 영상의 진위 여부를 학습한다. 여기서, 제1 감별자(254)는 생성자(220)와 연동하여 가상 영상 데이터가 참 신호에 해당하도록 분류하기 위하여 생성적 적대 신경망(GAN: Generative Adversarial Network)을 기반으로 학습을 수행하는 것이 바람직하나 반드시 이에 한정되는 것은 아니다.
- [0056] 제2 감별자(256)는 가상 영상 데이터 및 역방향 가상 영상 데이터에 대한 특징값을 이용하여 영상의 진행 방향에 대한 분류를 처리한다. 제2 감별자(256)는 가상 영상 데이터 및 역방향 가상 영상 데이터의 특징값을 비교하여 가상 영상의 진행 방향을 학습하고, 학습한 제2 학습 결과를 출력한다.
- [0057] 제2 감별자(256)는 제2 학습 결과에 근거하여 생성자(220)로 피드백 정보를 전달하며, 생성자(220)에서 순방향

가상 영상이 참 신호로 분류되는 가상 영상이 생성될 때까지 반복하여 가상 영상의 진위 여부를 학습한다. 여기서, 제2 감별자(256)는 생성자(220)와 연동하여 가상 영상 데이터가 참 신호에 해당하도록 분류하기 위하여 생성적 적대 신경망(GAN: Generative Adversarial Network)을 기반으로 학습을 수행하는 것이 바람직하나 반드시 이에 한정되는 것은 아니다.

- [0058] 감별자(250)는 제1 학습 결과 및 제2 학습 결과를 기반으로 피드백 정보를 산출하고, 피드백 정보를 이용하여 감별자(250)에서 가상 영상 데이터 및 실제 영상 데이터를 비교하여 순방향의 가상 영상이 참 신호로 분류될 때까지 반복하여 가상 영상의 진위 여부를 학습한다.
- [0059] 도 3은 본 발명의 실시예에 따른 영상 생성을 위한 학습 방법을 설명하기 위한 순서도이다.
- [0060] 영상 생성 장치(100)는 실제 영상 기반의 실제 영상 데이터를 획득한다(S310). 여기서, 실제 영상 데이터는 외부 장치로부터 획득될 수 있으며, 실제 영상을 기반으로 생성된 시간이 연속되는 적어도 하나의 세그먼트 블록을 포함할 수 있다.
- [0061] 영상 생성 장치(100)는 잠재 벡터를 획득하고(S320), 잠재 벡터를 기반으로 가상 영상을 생성하고, 생성된 가상 영상에 대한 가상 영상 데이터를 생성한다(S330). 영상 생성 장치(100)는 잠재 벡터를 기반으로 생성된 시간이 연속되는 적어도 하나의 세그먼트 블록을 포함하는 가상 영상 데이터를 생성한다. 여기서, 가상 영상 데이터에 포함된 상기 적어도 하나의 세그먼트 블록 각각은 RGB에 대한 특징값을 포함할 수 있다.
- [0062] 영상 생성 장치(100)는 가상 영상 데이터 및 실제 영상 데이터 각각의 재생 방향을 변환한다(S340). 영상 생성 장치(100)는 가상 영상 데이터 및 실제 영상 데이터 각각의 시간 특성을 반전시켜 역방향 가상 영상 데이터 및 역방향 실제 영상 데이터를 생성한다.
- [0063] 영상 생성 장치(100)는 제1 감별 처리를 통해 제1 학습결과를 생성한다(S350). 영상 생성 장치(100)는 가상 영상 데이터 및 실제 영상 데이터에 대한 특징값을 이용하여 상기 가상 영상의 진위 여부에 대한 분류를 처리한다. 영상 생성 장치(100)는 가상 영상 데이터 및 실제 영상 데이터의 특징값을 비교하여 가상 영상의 진위 여부를 학습하고, 학습한 제1 학습 결과를 출력한다.
- [0064] 영상 생성 장치(100)는 제2 감별 처리를 통해 제2 학습결과를 생성한다(S360). 영상 생성 장치(100)는 가상 영상 데이터 및 역방향 가상 영상 데이터에 대한 특징값을 이용하여 영상의 진행 방향에 대한 분류를 처리한다. 영상 생성 장치(100)는 가상 영상 데이터 및 역방향 가상 영상 데이터의 특징값을 비교하여 가상 영상의 진행 방향을 학습하고, 학습한 제2 학습 결과를 출력한다.
- [0065] 도 3에서는 각 단계를 순차적으로 실행하는 것으로 기재하고 있으나, 반드시 이에 한정되는 것은 아니다. 다시 말해, 도 3에 기재된 단계를 변경하여 실행하거나 하나 이상의 단계를 병렬적으로 실행하는 것으로 적용 가능할 것이므로, 도 3은 시계열적인 순서로 한정되는 것은 아니다.
- [0066] 도 3에 기재된 본 실시예에 따른 영상 생성 학습 방법은 애플리케이션(또는 프로그램)으로 구현되고 단말장치(또는 컴퓨터)로 읽을 수 있는 기록매체에 기록될 수 있다. 본 실시예에 따른 영상 생성 학습 방법을 구현하기 위한 애플리케이션(또는 프로그램)이 기록되고 단말장치(또는 컴퓨터)가 읽을 수 있는 기록매체는 컴퓨팅 시스템에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록장치 또는 매체를 포함한다.
- [0067] 도 4는 본 발명의 실시예에 따른 프로세서의 영상 생성을 위한 동작 구성을 개략적으로 나타낸 블록 구성도이다.
- [0068] 본 실시예에 따른 영상 생성 장치(100)에 포함된 프로세서(130)는 입력 벡터 획득부(410), 신경망 처리부(420), 학습 결과 적용부(430), 영상 생성부(440) 및 결과 출력부(450)를 포함한다. 도 4의 프로세서(130)는 일 실시예에 따른 것으로서, 도 4에 도시된 모든 블록이 필수 구성요소는 아니며, 다른 실시예에서 프로세서(130)에 포함된 일부 블록이 추가, 변경 또는 삭제될 수 있다. 한편, 프로세서(130)에 포함된 각 구성요소들은 각각 별도의 소프트웨어 장치로 구현되거나, 소프트웨어가 결합된 별도의 하드웨어 장치로 구현될 수 있다.
- [0069] 입력 벡터 획득부(410)는 영상 생성을 위한 입력 벡터를 획득한다. 여기서, 입력 벡터는 학습 시 입력되지 않은 영상을 생성하기 위한 입력값을 의미한다. 여기서, 입력 벡터는 랜덤으로 입력된 값 또는 사용자의 조작에 의해 입력된 값을 포함할 수 있다.
- [0070] 신경망 처리부(420)는 획득된 입력 벡터에 대한 특징값을 추출하는 동작을 수행한다. 신경망 처리부(420)는 컨볼루션 뉴럴 네트워크(CNN: Convolutional Neural Networks) 학습을 기반으로 특징값을 추출할 수 있다. 여기

서, 특징값은 복수의 영상 특징값을 포함할 수 있다.

- [0071] 학습 결과 적용부(430)는 가상 영상 데이터 및 실제 영상 데이터의 특징값을 비교하여 가상 영상의 진위 여부를 학습한 제1 학습 결과와 가상 영상 데이터 및 역방향 가상 영상 데이터의 특징값을 비교하여 가상 영상의 진행 방향을 학습한 제2 학습 결과를 적용하며, 영상 생성부(440)는 적용된 학습 결과를 기반으로 신규 영상을 생성한다. 여기서, 신규 영상은 가상 영상을 의미한다.
- [0072] 결과 출력부(450)는 생성된 신규 영상을 출력하여 제공한다. 여기서, 신규 영상은 영상 인식, 동작 인식 등의 학습을 위한 데이터 세트로 제공될 수 있다.
- [0073] 도 5는 본 발명의 실시예에 따른 영상 생성 방법을 설명하기 위한 순서도이다.
- [0074] 영상 생성 장치(100)는 영상 생성을 위한 입력 벡터를 획득한다(S510). 여기서, 입력 벡터는 학습 시 입력되지 않은 영상을 생성하기 위한 입력값을 의미한다. 여기서, 입력 벡터는 랜덤으로 입력된 값 또는 사용자의 조작에 의해 입력된 값을 포함할 수 있다.
- [0075] 영상 생성 장치(100)는 획득된 입력 벡터를 신경망 처리하여 특징값을 추출한다(S520). 여기서, 영상 생성 장치(100)는 컨볼루션 뉴럴 네트워크(CNN: Convolutional Neural Networks) 학습을 기반으로 입력 벡터에 대한 특징값을 추출할 수 있다.
- [0076] 영상 생성 장치(100)는 기 학습된 학습 결과 적용하여(S530), 신규 영상을 생성한다(S540).
- [0077] 영상 생성 장치(100)는 가상 영상 데이터 및 실제 영상 데이터의 특징값을 비교하여 가상 영상의 진위 여부를 학습한 제1 학습 결과와 가상 영상 데이터 및 역방향 가상 영상 데이터의 특징값을 비교하여 가상 영상의 진행 방향을 학습한 제2 학습 결과를 적용하며, 적용된 학습 결과를 기반으로 신규 영상을 생성한다. 여기서, 신규 영상은 가상 영상을 의미한다.
- [0078] 영상 생성 장치(100)는 생성된 신규 영상을 출력하여 제공한다. 여기서, 신규 영상은 영상 인식, 동작 인식 등의 학습을 위한 데이터 세트로 제공될 수 있다.
- [0079] 도 5에서는 각 단계를 순차적으로 실행하는 것으로 기재하고 있으나, 반드시 이에 한정되는 것은 아니다. 다시 말해, 도 5에 기재된 단계를 변경하여 실행하거나 하나 이상의 단계를 병렬적으로 실행하는 것으로 적용 가능할 것이므로, 도 5는 시계열적인 순서로 한정되는 것은 아니다.
- [0080] 도 5에 기재된 본 실시예에 따른 영상 생성 방법은 애플리케이션(또는 프로그램)으로 구현되고 단말장치(또는 컴퓨터)로 읽을 수 있는 기록매체에 기록될 수 있다. 본 실시예에 따른 영상 생성 방법을 구현하기 위한 애플리케이션(또는 프로그램)이 기록되고 단말장치(또는 컴퓨터)가 읽을 수 있는 기록매체는 컴퓨팅 시스템에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록장치 또는 매체를 포함한다.
- [0081] 도 6은 본 발명의 제1 실시예에 따른 영상 생성 장치의 학습 동작을 설명하기 위한 예시도이다.
- [0082] 본 발명의 제1 실시예에 따른 영상 생성 장치(100)의 프로세서(130)는 ArrowGAN 기반의 프레임워크를 적용하여 생성자(220)와 감별자(250)에서 인간처럼 영상에 대한 시간의 특성(AoT: Arrow of Time)을 인지하는 능력을 학습시키는 것을 목표로 한다.
- [0083] 이하, ArrowGAN 프레임워크에서의 감별자(250)에 대해 설명하도록 한다.
- [0084] 감별자(250)는 실제 영상과 페이크 영상(가상 영상)을 구별하고, 정방향 영상 및 역방향 영상도 구별한다.
- [0085] 감별자(250)는 입력 영상 x 가 실제 영상일 때의 확률 $p(\text{real} \mid x)$, 입력 영상 x 가 순방향일 때의 확률 $p(\text{forward} \mid x)$ 를 출력한다.
- [0086] 감별자(250)는 공유된 컨볼루션 레이어(252)와 두 개의 출력단(254, 256)으로 구성된다. 감별자(250)는 주어진 영상 x 에 대해, 자체 감독 이진 교차 엔트로피 손실을 설계하기 위해 수학적 1을 정의하고, 한 쌍의 순방향 영상 x^{forward} 및 반전을 통한 역방향 영상 x^{backward} 를 생성한다.

수학식 1

$$\mathcal{L}_D^{SS} = - \sum_{a \in A} \mathbb{E}_{x \sim P_{data}} [\log p(a|x^a)]$$

[0087]

[0088] 여기서, A는 AoT 세트 {forward, backward}를 의미하고, P_{data} 는 실제 영상의 분포를 의미하고, $p(a|x^a)$ 는 영상이 정방향 재생인지 역방향 재생인지에 대한 예측값의 베르누이 분포를 의미한다. 상기 목적을 최소화함으로써, 감별자(250)는 수학식 2를 이용하여 시간의 특성에 대한 감각을 학습한다.

수학식 2

$$\mathcal{L}^{adv} = \mathbb{E}_{x \sim P_{data}} [\log D(x)] + \mathbb{E}_{z \sim p_z} [\log(1 - D(G(z)))]$$

[0089]

[0090] 감별자(250)는 실제 영상일 때의 확률 $p(\text{real} | x)$ 학습에 대한 적대적 손실(L^{adv})을 유지하고, 실제 영상과 생성된 가상 영상을 구별하여 극대화하려고 한다. 감별자(250)의 완전한 목적함수는 수학식 3과 같이 정의될 수 있다.

수학식 3

$$\mathcal{L}_D = -\mathcal{L}^{adv} + \alpha \mathcal{L}_D^{SS}$$

[0091]

[0092] 여기서 α 는 두 항 사이의 중요도를 제어하기 위한 하이퍼변수이다.

[0093] 도 6에 도시된 바와 같이, 본 발명은 두 가지 방식으로 감별자(250)를 훈련시킨다. 실제 영상 및 생성된 가상 영상은 공유 컨볼루션 레이어와 L_D^{SS} 를 생성하는 제1 감별자(254)를 통과한다.

[0094] 또한, 순방향 영상 및 역방향 영상은 공유 컨볼루션 레이어와 L_D^{adv} 를 생성하는 제2 감별자(256)를 통과한다.

[0095] 이하, ArrowGAN 프레임워크에서의 생성자(220)에 대해 설명하도록 한다.

[0096] 생성자(220)는 제약 조건(inductive bias)을 기반으로 생성된 가상 영상이 시간 내에 순방향으로 진행되는 것을 학습한다. 따라서, 생성자(220)는 생성된 가상 영상 $G(z)$ 를 감별자(250)로 입력함으로써, 감별자(250)로부터 손실을 받는다. 여기서, 손실은 수학식 4와 같이 정의될 수 있다.

수학식 4

$$\mathcal{L}_G^{SS} = - \sum_{a \in A} \mathbb{E}_{z \sim p_z} [\log p(a|G(z)^a)]$$

[0097]

[0098] 여기서 p_z 는 표준 가우스 분포 $N(0, 1)$ 입니다. 위의 목적을 최소화함으로써, 생성자(220)는 순방향 영상을 생성하기 위한 학습을 수행한다. 생성자(220)는 순방향 영상만 제작한다. 하지만, 감별자(250)는 실제 영상과 유사하게 만들기 위해 순방향 영상과 순방향 영상을 반전시킨 역방향 영상을 모두를 입력 받는다.

[0099] 한편, 생성자(220)는 사실적으로 가상 영상을 생성하기 위하여 적대적 손실(\mathcal{L}^{adv})를 최소화한다.

수학식 5

$$\mathcal{L}_G = \mathcal{L}^{adv} + \beta \mathcal{L}_G^{SS}$$

[0100]

[0101] 생성자(220)는 생성자(220)에 대한 전체 목적을 최소화함으로써, 사실적이고 순방향으로 진행되는 영상을 생성하는 학습을 수행한다. 여기서 β 는 두 용어 사이의 중요도를 제어하기 위한 하이퍼변수이다.

[0102] 본 발명에 따른 ArrowGAN 프레임워크는 일반적으로 사용되고 있는 다양한 GAN 학습 방식에 적용될 수 있다. 예를 들어, 본 발명에 따른 ArrowGAN 프레임워크는 VGAN(video Generative Adversarial Nets), TGAN (Temporal Generative Adversarial Nets), MoCoGAN 등에 적용될 수 있다.

[0103] 도 7은 본 발명의 제2 실시예에 따른 영상 생성 장치의 학습 동작을 설명하기 위한 예시도이다.

[0104] 본 발명의 제2 실시예에 따른 영상 생성 장치(100)의 프로세서(130)는 범주형으로 ArrowGAN 기반의 프레임워크를 적용하여 영상을 생성하는 것을 목표로 한다. 다시 말해, 제2 실시예에 따른 영상 생성 장치(100)의 프로세서(130)는 이미지 생성에 최신 기술을 사용하여 범주형 MoCoGAN 베이스 라인(baseline)에 비해 성능을 개선하고 범주 영상 생성을 위한 ArrowGAN 프레임 워크를 활용한다.

[0105] 베이스 라인은 프레임 단위로 생성되며 두 개의 감별자가 존재한다. 하나는 영상용 감별자이고, 다른 하나는 프레임용 감별자이다.

[0106] 비디오 생성과 이미지 생성 간의 격차를 줄이기 위해 이미지 생성에서 범주 형 MoCo-GAN에 이르기까지 최근의 효과적인 기술을 사용할 수 있다.

[0107] 먼저, 조건부 배치 정규화 (CBN: Conditional Batch Normalization, 730)을 확장하여 클래스 레이블을 생성자(220)에 전달한다.

[0108] 조건부 배치 정규화 계층(730)은 클래스 레이블(720)의 출력 이미지를 조정하기 위해 배치 정규화를 위한 매개 변수를 조작하여 생성자의 기능 활성화를 변조한다. 여기서, 영상 생성을 위해 각 프레임 생성자(220)에 조건부 배치 정규화 레이어(730)를 삽입한다.

[0109] 또한, 감별자(250)는 프로젝션 감별자로 대체되며, 프로젝션 감별자는 두 번째 특징 벡터와 클래스 점수로 간주되는 클래스 임베딩 사이에서 내부 구성(inner product)을 계산한다. 3D 컨볼루션(740)을 사용하여 영상 도메인을 자연스럽게 일반화하기 위해 프로젝션 감별자를 수학식 6을 기반으로 수정할 수 있다.

수학식 6

$$D_{proj}(x, y) = y^T V \Phi(x) + \psi(\Phi(x))$$

[0110]

[0111] 여기서 y 는 클래스 레이블을 나타내는 one-hot 벡터이고, V 는 클래스 임베딩 행렬이고, Φ 는 감별자의 최종 계층에서 특징을 추출하며, \emptyset 는 스칼라 값을 생성하는 풀리 커넥티드 레이어를 의미한다.

[0112] 다음으로, 본 발명에서는 안정적인 훈련을 위해 비디오 감별자 및 프레임 감별자 모두에서 스펙트럼 정규화 계층을 활용한다. 이후, 각 영역에서 안정적인 훈련과 다양한 영상을 위해 모드 탐색 손실을 추가한다. 설정에서 주어진 2 개의 잠재 벡터에 대해 생성된 영상 간의 차이를 비디오 간의 차이를 증가시킨다. 마지막으로, 본 발명에서는 잠재 벡터 사이의 거리에 걸쳐 생성된 영상 사이의 거리를 명시적으로 증진함으로써 영상의 다양성을 강화하고 모드 붕괴 문제를 완화시킬 수 있다. 특히, 본 발명에서는 그 역수의 최소화를 수학식 7과 같이 수행할 수 있다.

수학식 7

$$\mathcal{L}_{MS} = \frac{d_z(Z_1, Z_2)}{d_v(G(Z_1, y), G(Z_2, y))}$$

여기서, d_z 와 d_v 는 잠재 벡터 사이의 거리와 생성된 영상 사이의 거리를 각각 나타낸다.

본 발명의 제2 실시예에 따른 영상 생성 장치(100)의 프로세서(130)는 적대적 손실이 힙지 버전(hinge version of the adversarial losses), 자체 감독 손실(self-supervised losses) 및 모드 탐색 정규화기(mode-seeking regularizer) 등의 목적 함수로 범주형 ArrowGAN를 훈련한다. 생성자와 감별자의 훈련은 수학식 8과 같이, 주어진 최대-최소 문제를 해결함으로써 달성된다.

수학식 8

$$\begin{aligned}\mathcal{L}_{D_v}^{adv} &= -\mathbb{E}_{x,y \sim P_{data}} [\min(0, 1 + D_{proj}(x, y))] \\ &\quad - \mathbb{E}_{\tilde{x} \sim P_G, y \sim P_{data}} [\min(0, 1 - D_{proj}(\tilde{x}, y))] \\ \mathcal{L}_{G_v}^{adv} &= -\mathbb{E}_{\tilde{x} \sim P_G, y \sim P_{data}} [D(\tilde{x}, y)] \\ \mathcal{L}_{img}^{adv} &= \mathbb{E}_{x_i \sim P_{data}} [\log D_i(x_i)] \\ &\quad + \mathbb{E}_{\tilde{x}_i \sim P_G} [\log(1 - D_i(\tilde{x}_i))]\end{aligned}$$

마지막으로, 생성자와 감별자의 전체 목적은 수학식 9와 같다.

수학식 9

$$\begin{aligned}\mathcal{L}_{D_{video}} &= \mathcal{L}_{D_v}^{adv} + \lambda_1 L_D^{SS} \\ \mathcal{L}_{D_{image}} &= -\mathcal{L}_{img}^{adv} \\ \mathcal{L}_G &= \mathcal{L}_{G_v}^{adv} + \mathcal{L}_{img}^{adv} + \lambda_2 L_G^{SS} + \lambda_3 L_{MS}\end{aligned}$$

여기서, λ_1 , λ_2 , 및 λ_3 은 1, 0.2 및 0.2 각각으로 가정할 수 있다. 범주형 ArrowGAN에 대한 구조는 도 7과 같다. 범주형 ArrowGAN은 클래스 레이블(720)의 입력을 통해 레이블 정보에 대응하는 가상 영상(제1 내지 제4 가상 영상)을 생성하고, 이를 통해 생성적 적대 학습을 수행할 수 있다.

도 8a 내지 도 8c는 본 발명의 실시예에 따른 영상 생성 장치의 학습 결과 및 적용 결과를 나타낸 도면이다.

도 8a는 범주형 ArrowGAN에서 생성된 선별되지 않은 영상 세트를 나타낸다. 도 8a는 Weizmann, UCFsports 및 UCF-101에서 클래스 레이블별로 유효한 비디오를 정 성적으로 얼마나 잘 생성하는지를 보여준다.

도 8a의 (a, b)는 서로 다른 8 개의 클래스에 대한 4 개의 연속 프레임을 나타내고, 도 8a의 (c)는 많은 비디오에서 무작위로 샘플링 된 프레임을 나타낸다.

도 8a의 (a, b)는 클래스 레이블만 있는 잠재 공간에서 생성된 샘플이다. 또한, 범주형 ArrowGAN은 도 8a의 (c)와 같이 단순한 데이터 세트뿐만 아니라 대규모 데이터 세트 UCF-101에서도 영상을 생성할 수 있다.

도 8b는 일반적인 GAN 학습의 베이스라인에 ArrowGAN 프레임워크를 적용한 효과를 나타낸다.

도 8b는 ArrowGAN이 적용될 때 모든 베이스라인과 모든 데이터 세트에 대해 IS의 지속적인 개선을 정량적으로

보여준다. 본 발명에서는 생성자를 수정하지 않고, 감별자에 최소한으로 변경하여 보조 자체 감독 작업을 추가한다. 이는 ArrorGAN 프레임워크가 다른 영상-GAN이에 쉽게 적용할 수 있음을 의미한다.

[0126] 도 8c는 ArrowGAN의 정성적 결과를 나타낸다. 도 8c의 (a)는 종래의 GAN 학습 결과를 나타내고, 도 8c의 (b)는 ArrorGAN 프레임워크를 적용한 GAN 학습 결과를 나타낸다. 도 8c의 (b)에서는 특히, 사지(four limbs) 또는 사물과 같은 세밀한 부분에서 개선된 것을 관찰할 수 있다.

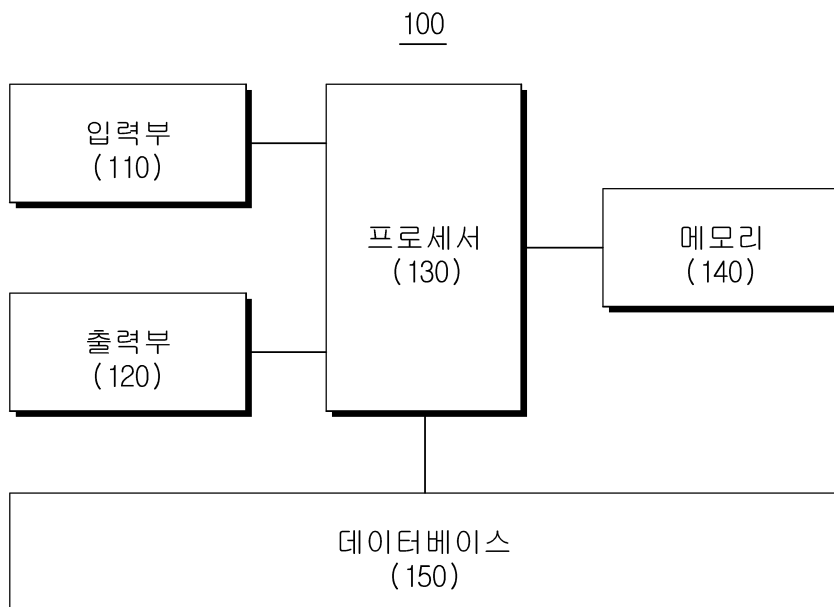
[0127] 이상의 설명은 본 발명의 실시예의 기술 사상을 예시적으로 설명한 것에 불과한 것으로서, 본 발명의 실시예가 속하는 기술 분야에서 통상의 지식을 가진 자라면 본 발명의 실시예의 본질적인 특성에서 벗어나지 않는 범위에서 다양한 수정 및 변형이 가능할 것이다. 따라서, 본 발명의 실시예들은 본 발명의 실시예의 기술 사상을 한정하기 위한 것이 아니라 설명하기 위한 것이고, 이러한 실시예에 의하여 본 발명의 실시예의 기술 사상의 범위가 한정되는 것은 아니다. 본 발명의 실시예의 보호 범위는 아래의 청구범위에 의하여 해석되어야 하며, 그와 동등한 범위 내에 있는 모든 기술 사상은 본 발명의 실시예의 권리범위에 포함되는 것으로 해석되어야 할 것이다.

부호의 설명

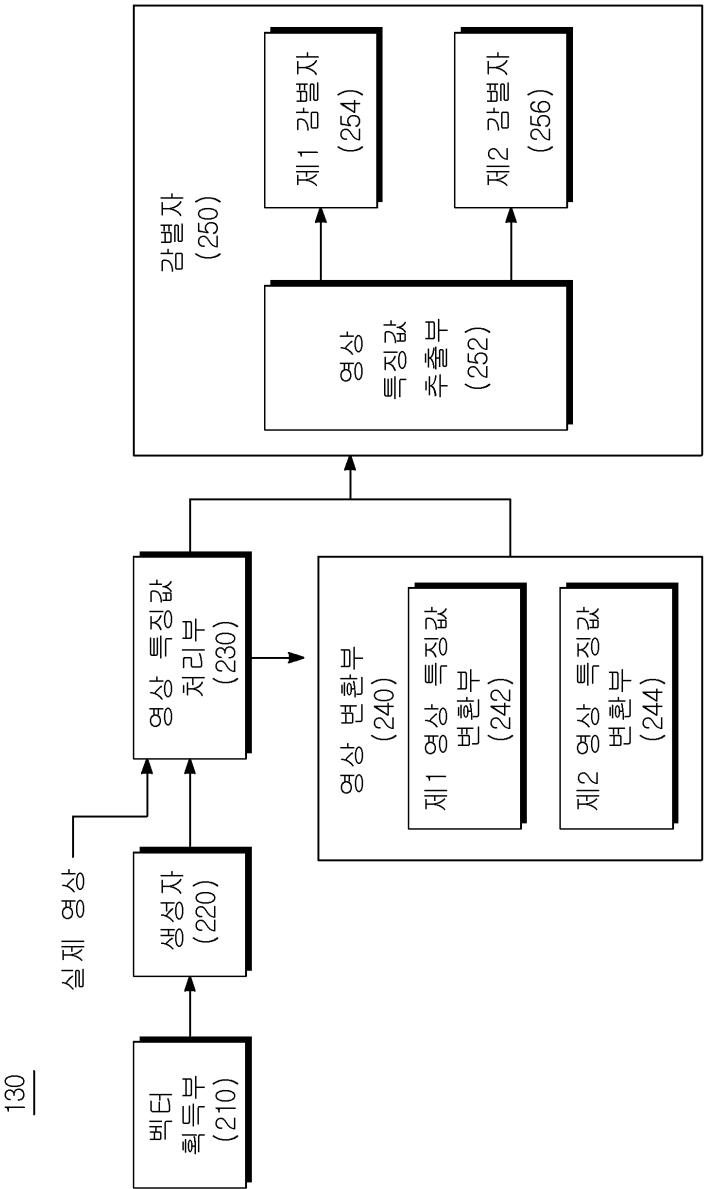
[0128] 100: 영상 생성 장치
110: 입력부
120: 출력부
130: 프로세서
140: 메모리
150: 데이터 베이스
210: 벡터 획득부
220: 생성자
230: 영상 특징값 처리부
240: 영상 변환부
250: 감별자

도면

도면1

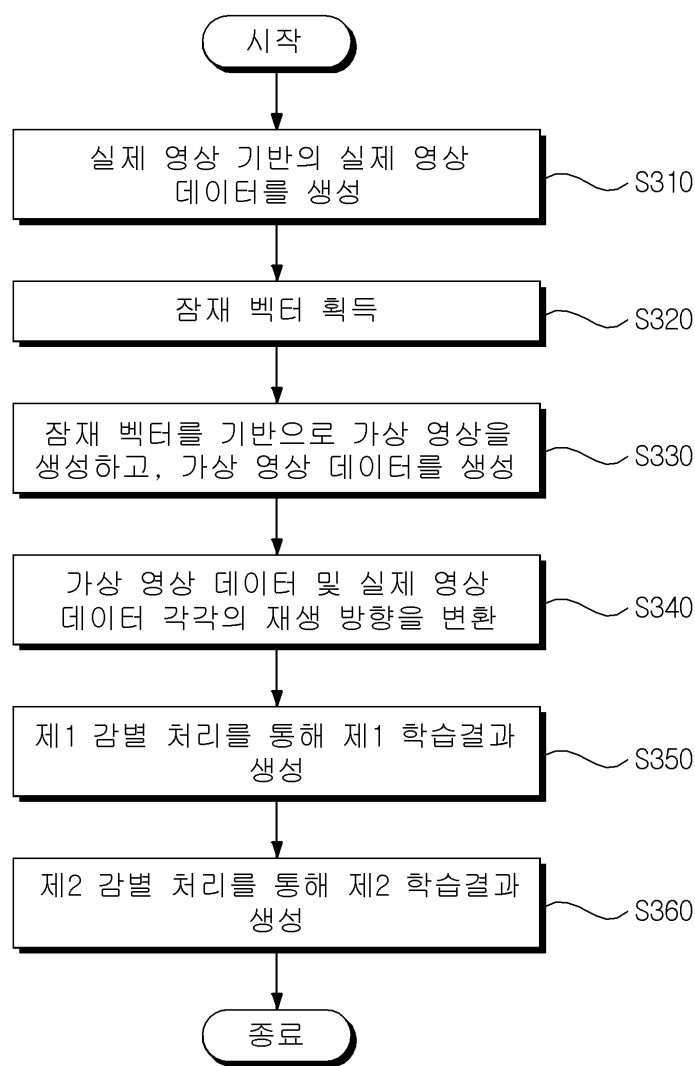


도면2

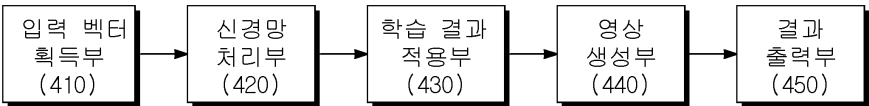


130

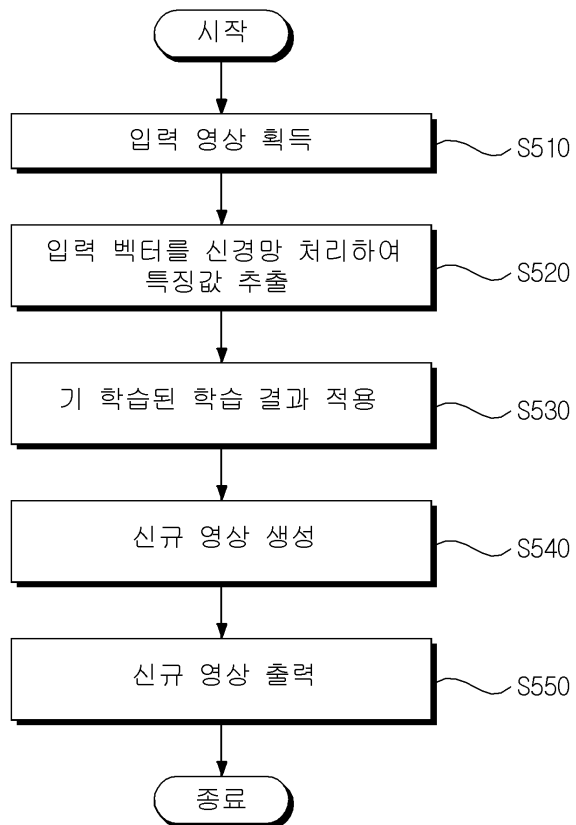
도면3



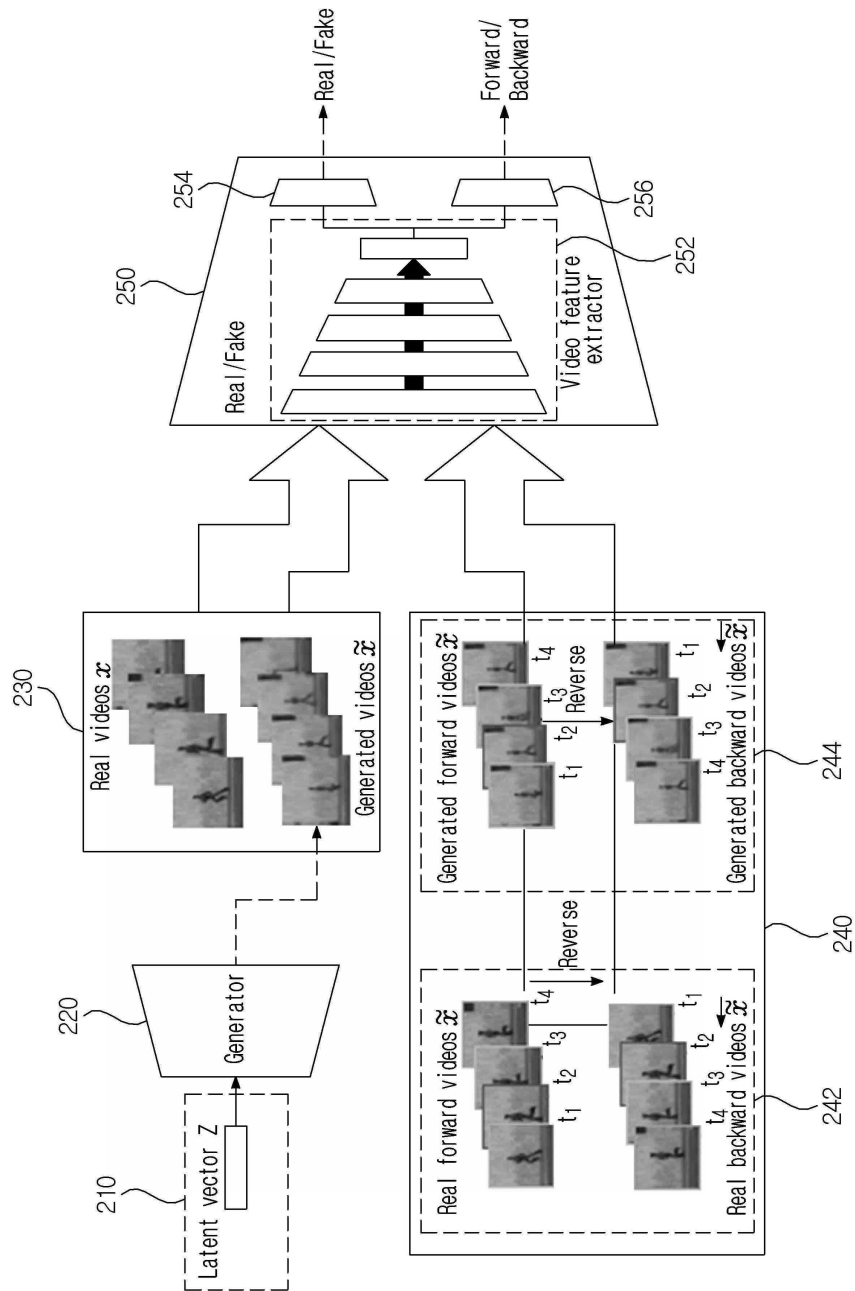
도면4



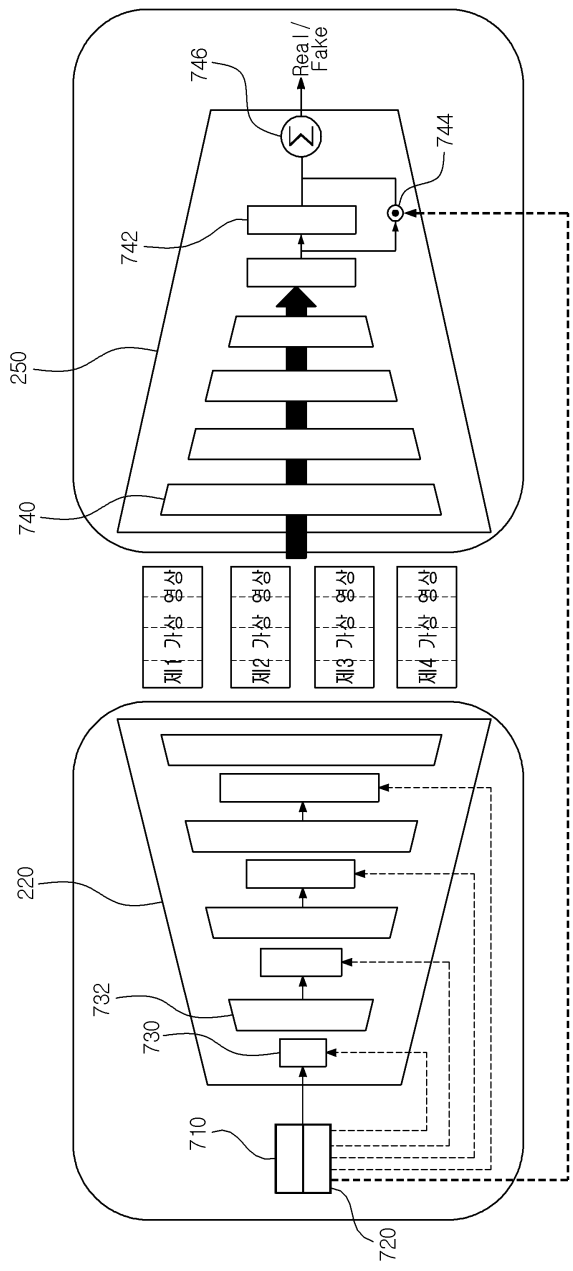
도면5



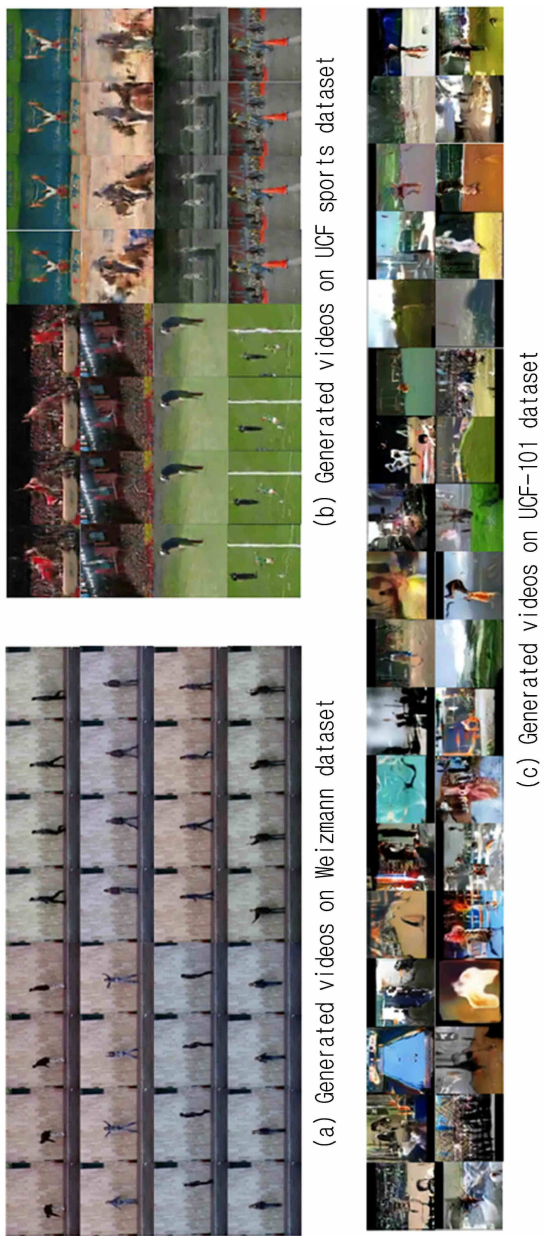
도면6



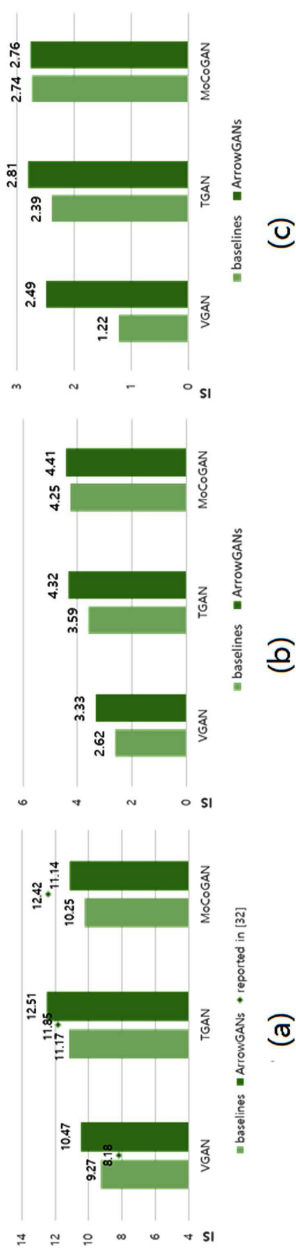
도면7



도면8a



도면8b



도면8c

