(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2020년12월16일
(11) 등록번호 10-2192209
(24) 등록일자 2020년12월10일

(51) 국제특허분류(Int. Cl.)
G06T 7/246 (2017.01) *G06N 3/08* (2006.01)
G06T 5/00 (2019.01)
(52) CPC특허분류
G06T 7/251 (2017.01)
G06N 3/08 (2013.01)
(21) 출원번호 10-2020-0076548
(22) 출원일자 2020년06월23일
심사청구일자 2020년06월23일
(56) 선행기술조사문헌
Haoye Cai외 3인, "Deep video generation, predicion and completion of human action sequences", ECCV 2018,(2018.09.14.), 1부.*
Jason Kugelman외 5인, "Constructing synthetic chorio_retinal patches using generative adversarial networks", DICTA2019, (2019.12.04.), 1부.*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
인하대학교 산학협력단
인천광역시 미추홀구 인하로 100(용현동, 인하대학교)
(72) 발명자
조근식
인천광역시 연수구 아카데미로312번길 177, 104동 2903호(송도동, 힐스테이트 레이크 송도 1차)
아지즈
인천광역시 미추홀구 인하로 100(용현동, 인하대학교)
(74) 대리인
양성보

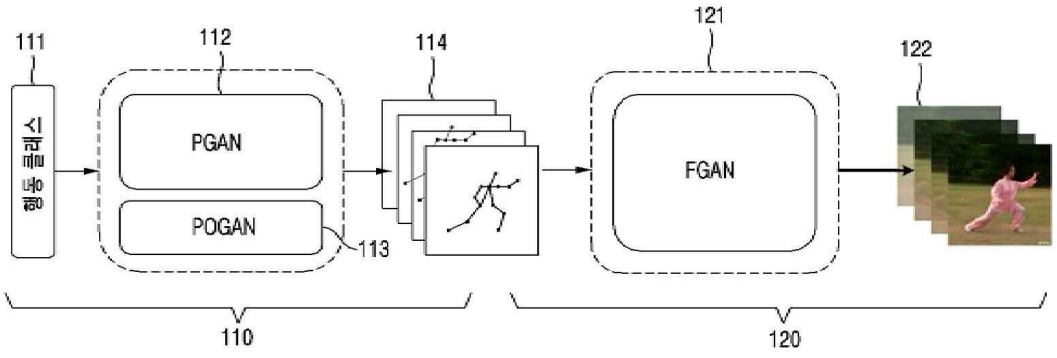
전체 청구항 수 : 총 4 항 심사관 : 박금옥

(54) 발명의 명칭 **인간 행동 생성을 위한 GAN 조직화 방법 및 장치**

(57) 요 약

인간 행동 생성을 위한 GAN 조직화 방법 및 장치가 제시된다. 본 발명에서 제안하는 인간 행동 생성을 위한 GAN 조직화 방법은 행동 표현부를 통해 골격 포인트의 변화 패턴을 구성하는 인체 움직임을 행동 표현으로 나타내는 단계, 포즈 생성부를 통해 행동 표현의 클래스 등급을 입력 받아 행동 표현을 재구성하고 판별하여 포즈를 생성하는 포즈 생성 단계, 생성된 포즈의 서로 다른 노이즈 레벨 입력과 예상 스타일 출력을 포즈 최적화부를 통해 쌍으로 매핑하여 노이즈를 제거하고, 순차적인 골격 포즈를 생성하는 포즈 최적화 단계 및 노이즈가 제거된 포즈를 입력 받아 프레임 생성부를 통해 프레임 시퀀스로 변환하는 프레임 생성 단계를 포함한다.

대 표 도



(52) CPC특허분류
G06T 5/002 (2013.01)
G06T 2207/20084 (2013.01)

이 발명을 지원한 국가연구개발사업
과제고유번호 1711103050
과제번호 2017-0-01642-004
부처명 과학기술정보통신부
과제관리(전문)기관명 정보통신기획평가원
연구사업명 국책과제 운영지원비(산학회계)
연구과제명 [간접비]인공지능을 활용한 콘텐츠 창작 기술-센터운영비(4차년도)
기 여 율 1/1
과제수행기관명 인하대학교 산학협력단
연구기간 2020.01.01 ~ 2020.12.31
공지예외적용 : 있음

명 세 서

청구범위

청구항 1

행동 표현부를 통해 골격 포인트의 변화 패턴을 구성하는 인체 움직임을 복수의 프레임 내에서 골격 포즈의 각 지점의 변화를 나타내는 행동 표현으로 나타내는 단계;

포즈 생성부를 통해 행동 표현의 클래스 등급을 입력 받아 행동 표현을 재구성하고 판별하여 포즈를 생성하는 포즈 생성 단계;

생성된 포즈의 서로 다른 노이즈 레벨 입력과 예상 스타일 출력을 포즈 최적화부를 통해 쌍으로 매핑하여 노이즈를 제거하고, 순차적인 골격 포즈를 생성하는 포즈 최적화 단계; 및

노이즈가 제거된 순차적인 골격 포즈를 입력 받아 프레임 생성부를 통해 프레임 시퀀스로 변환하는 프레임 생성 단계

를 포함하고,

포즈 생성부를 통해 행동 표현의 클래스 등급을 입력 받아 행동 표현을 재구성하고 판별하여 포즈를 생성하는 포즈 생성 단계는,

생성기 및 판별기를 포함하는 포즈 생성부의 생성기를 통해 잠재 노이즈에 대한 밀집 계층, 배치 정규화, ReLU를 수행한 후 클래스 등급과의 결합을 위한 동일한 차원으로 재구성하고, 클래스 등급에 대한 임베딩, 밀집 계층을 수행한 후 잠재 차원 노이즈와의 결합을 위한 동일한 차원으로 재구성한 후 잠재 차원 노이즈와 클래스 등급을 결합하고, 컨볼루션 블록 계층을 수행하여 포즈 생성부의 판별기로 전송하고,

포즈 생성부의 판별기를 통해 포즈 생성부의 생성기와 같은 클래스 등급에 대한 임베딩, 밀집 계층을 수행한 후 포즈 생성부의 생성기로부터 입력 받은 행동 표현과의 결합을 위한 동일한 차원으로 재구성한 후 행동 표현과 클래스 등급을 결합하고, 컨볼루션 블록 계층을 수행하여 포즈를 판별하는

GAN(Generative Adversarial Networks) 조직화 방법.

청구항 2

삭제

청구항 3

삭제

청구항 4

제1항에 있어서,

노이즈가 제거된 순차적인 골격 포즈를 입력 받아 프레임 생성부를 통해 프레임 시퀀스로 변환하는 프레임 생성 단계는,

골격 이미지를 입력으로 하여 골격의 텍스처를 오버레이하고, 장면 프레임을 생성하는

GAN 조직화 방법.

청구항 5

골격 포인트의 변화 패턴을 구성하는 인체 움직임을 복수의 프레임 내에서 골격 포즈의 각 지점의 변화를 나타내는 행동 표현으로 나타내는 행동 표현부;

행동 표현의 클래스 등급을 입력 받아 행동 표현을 재구성하고 판별하여 포즈를 생성하는 포즈 생성부;

생성된 포즈의 서로 다른 노이즈 레벨 입력과 예상 스타일 출력을 쌍으로 매핑하여 노이즈를 제거하고, 순차적

인 골격 포즈를 생성하는 포즈 최적화부; 및

노이즈가 제거된 순차적인 골격 포즈를 입력 받아 프레임 시퀀스로 변환하는 프레임 생성부

를 포함하고,

포즈 생성부는,

생성기 및 판별기를 포함하고, 포즈 생성부의 생성기를 통해 잠재 노이즈에 대한 밀집 계층, 배치 정규화, ReLU를 수행한 후 클래스 등급과의 결합을 위한 동일한 차원으로 재구성하고, 클래스 등급에 대한 임베딩, 밀집 계층을 수행한 후 잠재 차원 노이즈와의 결합을 위한 동일한 차원으로 재구성한 후 잠재 차원 노이즈와 클래스 등급을 결합하고, 컨볼루션 블록 계층을 수행하여 포즈 생성부의 판별기로 전송하고,

포즈 생성부의 판별기를 통해 포즈 생성부의 생성기와 같은 클래스 등급에 대한 임베딩, 밀집 계층을 수행한 후 포즈 생성부의 생성기로부터 입력 받은 행동 표현과의 결합을 위한 동일한 차원으로 재구성한 후 행동 표현과 클래스 등급을 결합하고, 컨볼루션 블록 계층을 수행하여 포즈를 판별하는

GAN(Generative Adversarial Networks) 조직화 장치.

청구항 6

삭제

청구항 7

삭제

청구항 8

제5항에 있어서,

프레임 생성부는,

골격 이미지를 입력으로 하여 골격의 텍스처를 오버레이하고, 장면 프레임을 생성하는

GAN 조직화 장치.

발명의 설명

기 술 분 야

[0001] 본 발명은 인간 행동 생성을 위한 GAN(Generative Adversarial Networks) 조직화 방법 및 장치에 관한 것이다.

배 경 기 술

[0002] 종래기술1(Goodfellow, I., et al.: Generative adversarial nets. In: NIPS (2014))의 생성적 대립 신경망(Generative Adversarial Networks; GAN)은 이미지 및 비디오 생성과 같은 콘텐츠 생성 과제에서 빠른 발전을 보였다. 그 중에서도 인간과 객체 애니메이션을 위한 다양한 어플리케이션의 발달로 인한 특별한 관심이 비디오 생성을 가져온다. 비디오 생성 과제는 이미지 생성과 달리 특별한 접근이 필요하다. 비디오가 함께 결합된 프레임 이미지의 집합체라 하더라도 장면 역학, 이전 및 종합적인 움직임과의 관계, 콘텐츠의 일관성 등 생성 과제 달성을 위해 고려해야 할 요소가 다양하다. 이러한 분야는 심층적인 연구와 분석이 필요하지만 이미 주목할 만한 결과가 제시됐다.

[0003] 다양한 종래기술에서, 비디오 생성 방법을 제공하려고 시도하였다.

[0004] 종래기술2(Saito, M., Matsumoto, E., Saito, S.: Temporal generative adversarial nets with singular value clipping. In: ICCV (2017))에서는 잠재 변수 생성과 그 변수의 비디오 변환을 사용했다. 또 다른 종래기술3(Marwah, T., Mittal, G., Balasubramanian, V.N.: Attentive semantic video generation using captions. In: 2017 IEEE International Conference on Computer Vision (ICCV), pp. 1435-1443. IEEE (2017)) 및 종래기술4(Vondrick, C., Pirsiavash, H., Torralba, A.: Generating videos with scene dynamics. In: Lee, D.D., Sugiyama, M., Luxburg, U.V., Guyon, I., Garnett, R. (eds.) Advances in Neural Information Processing Systems 29, pp. 613-621. Curran Associates, Inc. (2016))은 비디오의 직접 픽셀을 생성하려고

시도했다. 종래기술5(Tulyakov, S., Liu, M.Y., Yang, X., Kautz, J.: MoCoGAN: decomposing motion and content for video generation. arXiv preprint arXiv:1707.04993 (2017))은 임의 벡터에서 일련의 프레임을 생성하였다. 그러나 인간 행동의 발생에 초점을 맞춘 종래기술6(Yang, C., Wang, Z., Zhu, X., Huang, C., Shi, J., Lin, D.: Pose guided human video generation. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11214, pp. 204-219. Springer, Cham (2018)) 및 종래기술7(Cai, H., Bai, C., Tai, Y.-W., Tang, C.-K.: Deep video generation, prediction and completion of human action sequences. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11206, pp. 374-390. Springer, Cham (2018))에서는 비디오 역학은 인간 포즈로 제어하고 관리해야 한다는 것을 보여주었다. 이러한 종래기술은 둘 다 비디오 프레임 생성의 유사한 파이프라인을 적용하였다. 첫째, 일련의 포즈들을 생산하고 다음으로 장면을 만들기 위해 포즈들을 사용한다. 두 작업 모두 나름대로의 방법을 제시했다. 종래기술6에서, 움직임의 초기 포즈를 입력으로 취하고 프레임 작성을 위해 다음 GAN에 공급할 일련의 포즈를 생성하는 GAN을 도입하였다.

[0005] 반대로 포즈 생성 단계를 위해 종래기술7에서는 인간 골격 움직임 생성을 위해 함께 작동하는 두 개의 GAN을 적용했다. 한 GAN은 노이즈를 포즈로 매핑하고, 동시에 다른 GAN은 노이즈를 발생시켜 첫 번째 노이즈를 재사용한다. 그러나 우리는 인간 행동 생성을 위한 또 다른 방법을 제안했다.

[0006] 종래기술1에서 GAN의 도입은 비디오 생성 과제에서 다양한 연구 작업의 개발을 촉발시켰고, 이를 위한 많은 접근법이 있어 왔다. 종래기술2는 비디오의 의미표현을 학습하여 자체 제작할 수 있는 GAN을 제안했다. 네트워크는 나중에 동영상의 프레임에 매핑되는 일련의 잠재 변수를 생성하며, 이것은 나중에 하나의 비디오로 결합된다. 종래기술3은 자막을 이용한 무감독 방식의 비디오 생성법을 제안했다. 종래기술3의 연구자들은 단기, 장기의 시간적, 공간적 맥락을 이용하여 프레임 간의 일관성을 보존하는 비디오를 생성할 수 있다. 종래기술5는 비디오 생성 GAN에서 내용 및 동작의 분해 접근법을 사용했다. 동작과 내용 부분을 포함하는 일련의 임의 벡터를 사용하여 종래기술5에서 제안된 프레임워크는 벡터 대 프레임 매핑을 수행한다. 특정 행동에 따라 동작 부분은 바뀌는 반면, 내용 부분은 고정된 상태로 유지된다. 언급된 작업들은 유망한 결과를 보여주었지만, 종래기술6과 종래기술7에서 제안한 2단계 휴먼 비디오 생성은 포즈 제어 가능한 장면 생성(텍스처 골격 포즈가 만든 최종 프레임)이 안정적이고 더 그럴듯한 결과를 만들어 낸다는 것을 보여주었다.

[0007] 종래기술6은 포즈 생성을 위해 GAN을 적용하였는데, 초기 포즈와 행동등급을 입력으로 하고 포즈 시퀀스를 인코더-디코더 방식으로 합성한다. 여러 개의 컨볼루션 계층을 사용하여 인코딩한 포즈 시퀀스에는 행동 클래스 등급이 내장되어 있고, 몇 개의 잔여 블록 후에 결과는 일련의 부분 스트라이핑된 공간 컨볼루션 계층으로 디코더 부분을 통과했다. 또한 안정적이고 더 나은 학습을 위해 포즈 생성기 네트워크는 통합된 LSTM 모듈을 적용했다. 프레임 생성 GAN은 네트워크를 인간 포즈와 그에 상응하는 입력 이미지로 조절함으로써 이루어진다.

[0008] 종래기술7은 포즈 생성 단계를 위해 이중 GAN 아키텍처를 적용했다. 첫 번째 GAN은 행동 등급을 취하여 무작위 노이즈로 연결한 후 하나의 포즈를 연출한다. 동시에, 두 번째 GAN은 "적절한" 잠재 무작위 벡터 생성을 담당하며 이는 첫 번째 GAN에 의해 사용된다. 따라서 두 번째 GAN은 첫 번째 GAN을 여러 번 재사용하여 일련의 포즈들을 만들어 낸다.

발명의 내용

해결하려는 과제

[0009] 본 발명이 이루고자 하는 기술적 과제는 세 개의 GAN의 협업을 적용하여 첫 번째 네트워크-PGAN(Pose generation GAN)은 포즈 표현을 처리하고, 두 번째 네트워크-POGAN(Pose Optimization GAN)은 PGAN의 결과를 최적화하며, 세 번째 네트워크-FGAN(Frame generation GAN)은 골격 그림을 이미지 프레임으로 변환함으로써 움직임을 위한 특별한 데이터 표현을 사용하여 인간 행동 비디오를 생성하는 방법 및 장치를 제공하는데 있다.

과제의 해결 수단

[0010] 일 측면에 있어서, 본 발명에서 제안하는 인간 행동 생성을 위한 GAN 조직화 방법은 행동 표현부를 통해 골격 포인트의 변화 패턴을 구성하는 인체 움직임을 행동 표현으로 나타내는 단계, 포즈 생성부를 통해 행동 표현의 클래스 등급을 입력 받아 행동 표현을 재구성하고 판별하여 포즈를 생성하는 포즈 생성 단계, 생성된 포즈의 서로 다른 노이즈 레벨 입력과 예상 스타일 출력을 포즈 최적화부를 통해 쌍으로 매핑하여 노이즈를 제거하고, 순차적인 골격 포즈를 생성하는 포즈 최적화 단계 및 노이즈가 제거된 포즈를 입력 받아 프레임 생성부를 통해 프

레임 시퀀스로 변환하는 프레임 생성 단계를 포함한다.

[0011] 포즈 생성부를 통해 행동 표현의 클래스 등급을 입력 받아 행동 표현을 재구성하고 판별하여 포즈를 생성하는 포즈 생성 단계는 포즈 생성부의 생성기를 통해 행동 표현의 클래스 등급을 입력 받아 정규적으로 분산된 잠재 차원 노이즈를 획득하며, 임베딩, 클래스 등급에 대한 밀집 계층, 배치 정규화, 잠재 노이즈에 대한 ReLU를 수행한 후 결합을 위해 동일한 차원으로 재구성하여 포즈 생성부의 판별기로 전송한다.

[0012] 포즈 생성부를 통해 행동 표현의 클래스 등급을 입력 받아 행동 표현을 재구성하고 판별하여 포즈를 생성하는 포즈 생성 단계는 포즈 생성부의 판별기를 통해 생성기와 같은 클래스 등급으로 동일한 단계를 수행하여 생산된 출력과 결합하고, 컨볼루션 블록 계층을 수행하여 판별함으로써 포즈를 생성한다.

[0013] 노이즈가 제거된 포즈를 입력 받아 프레임 생성부를 통해 프레임 시퀀스로 변환하는 프레임 생성 단계는 골격 이미지를 입력으로 하여 골격의 텍스처를 오버레이하고, 장면 프레임을 생성한다.

[0014] 또 다른 일 측면에 있어서, 본 발명에서 제안하는 인간 행동 생성을 위한 GAN 조직화 장치는 골격 포인트의 변화 패턴을 구성하는 인체 움직임을 행동 표현으로 나타내는 행동 표현부, 행동 표현의 클래스 등급을 입력 받아 행동 표현을 재구성하고 판별하여 포즈를 생성하는 포즈 생성부, 생성된 포즈의 서로 다른 노이즈 레벨 입력과 예상 스타일 출력을 쌍으로 매핑하여 노이즈를 제거하고, 순차적인 골격 포즈를 생성하는 포즈 최적화부 및 노이즈가 제거된 포즈를 입력 받아 프레임 시퀀스로 변환하는 프레임 생성부를 포함한다.

발명의 효과

[0015] 본 발명의 실시예들에 따르면 세 개의 GAN의 협업을 적용하여 첫 번째 네트워크-PGAN(Pose generation GAN)은 포즈 표현을 처리하고, 두 번째 네트워크-POGAN(Pose Optimization GAN)은 PGAN의 결과를 최적화하며, 세 번째 네트워크-FGAN(Frame generation GAN)은 골격 그림을 이미지 프레임으로 변환함으로써 움직임을 위한 특별한 데이터 표현을 사용하여 인간 행동 비디오를 생성할 수 있다.

도면의 간단한 설명

[0016] 도 1은 본 발명의 일 실시예에 따른 인간 행동 생성을 위한 GAN 조직화 과정을 설명하기 위한 개략도 이다.

도 2는 본 발명의 일 실시예에 따른 인간 행동 생성을 위한 GAN 조직화 방법을 설명하기 위한 흐름도이다.

도 3은 본 발명의 일 실시예에 따른 행동 표현의 샘플을 나타내는 도면이다.

도 4는 본 발명의 일 실시예에 따른 인간 행동 생성을 위한 GAN 조직화 장치의 구성을 나타내는 도면이다.

도 5는 본 발명의 일 실시예에 따른 PGAN의 아키텍처를 설명하기 위한 도면이다.

도 6은 본 발명의 일 실시예에 따른 컨볼루션 블록을 설명하기 위한 도면이다.

도 7은 본 발명의 일 실시예에 따른 POGAN의 출력을 비교한 결과이다.

도 8은 본 발명의 일 실시예에 따른 포즈 생성 상태의 샘플 결과를 나타내는 도면이다.

발명을 실시하기 위한 구체적인 내용

[0017] 본 발명에서는, 인간 행동 생성을 위한 방법, 인간 행동 생성을 위한 GAN 조직화(GANs orchestraion for Human Actions Generation; GOHAG)를 제안한다. 움직임을 위한 특별한 데이터 표현을 사용하여 인간 행동 비디오 생성 작업에 세 개의 GAN의 협업을 적용한다. 첫 번째 네트워크-PGAN(Pose generation GAN)은 포즈 표현을 처리하고, 두 번째 네트워크-POGAN(Pose Optimization GAN)은 PGAN의 결과를 최적화하며, 세 번째 네트워크-FGAN(Frame generation GAN)은 골격 그림을 이미지 프레임으로 변환한다. 기존 작업과 비교했을 때, GOHAG는 생성된 이미지가 특정 운동에 대한 정보를 요약하는 이미지 합성으로 이 과제를 고려함으로써 포즈 생성을 위한 새로운 관점을 제공한다. 인간 행동 생성을 위해 GAN의 GOHAG 조직화를 적용하면, 초기 포즈를 제공하지 않고도 인간 생동을 생성할 수 있다. 더욱이, GOHAG는 특정한 행동을 나타내는 임의 벡터를 학습하는 대신에 움직임의 패턴을 생성하는 법을 학습한다. 이하, 본 발명의 실시 예를 첨부된 도면을 참조하여 상세하게 설명한다.

[0019] 도 1은 본 발명의 일 실시예에 따른 인간 행동 생성을 위한 GAN 조직화 과정을 설명하기 위한 개략도 이다.

[0020] 생성적 대립 신경망(Generative Adversarisl Networks; GAN)은 콘텐츠 창조 기술 개발에 큰 기여를 했다. 이러한 발전에서 중요한 부분은 휴먼 애니메이션 어플리케이션, 자동 트레일러 또는 영화 제작 등의 필요성으로 비

디오 생성을 한다는 점이다. 따라서 다양한 GAN의 장점을 활용하여 인체 움직임의 비디오를 생성하는 독자적인 방법 인간 행동 생성을 위한 GAN 조직화(GANs Orchestratio for Human Actions Generations; GOHAG)를 제안한다. GOHAG는 3개의 GAN으로 구성된다. 여기서 포즈 생성 GAN(Poses generation GAN; PGAN)은 포즈의 시퀀스를 생성하고, 포즈 최적화 GAN(Poses Optimization GAN; POGAN)이 최적화하며, 프레임 생성 GAN(Frames generation GAN; FGAN)은 시퀀스를 위한 텍스처를 추가하여 영상을 만든다. 제안된 방법은 매끄럽고 그럴듯한 고화질의 비디오를 생성하며 모델 기술 중에서 잠재력을 보여주었다.

[0021] 텍스처 부분의 경우, 본 발명은 참조 포즈 이미지를 촬영하고 인간 이미지를 출력하는 스킵 연결이 있는 컨볼루션 오토엔코더로 U-Net과 같은 네트워크를 적용했다. 본 발명에서는 종래기술6 및 종래기술7과 같이 2단계의 포즈 유도 비디오 생성을 적용하였지만, 포즈 생성을 위한 고유한 데이터 표현 방식과 자체 제작 방식을 제안하였다.

[0022] GOHAG 방법은 포즈 생성 및 최적화 단계(110) 및 프레임 생성 단계(120)를 포함한다. 먼저, 포즈 생성 및 최적화 단계(110)에서 특별한 방법으로 인간의 행동을 나타낸다. 도 1을 참조하면, 공간에서 정확한 골격 지점의 생성에 초점을 맞추는 대신 골격 움직임을 바탕으로 한 특정한 인간 행동의 패턴을 학습하여 행동 클래스를 생성한다(111). 다음으로, 인간 행동 생성을 위한 2단계 GAN 조정(112, 113)을 제안한다. 1단계(112)에서는 사용자 정의 행동 클래스를 입력으로 삼고 포즈 시퀀스 생성을 위해 POGAN과 함께 작동하는 제안된 PGAN에 제공한다. PGAN은 포즈 패턴을 생성하지만 POGAN은 이를 최적화하여 노이즈를 줄인다. GOHAG의 2단계는 텍스처 생성에 집중한다. 프레임 생성 단계(120)에서 FGAN(121)은 생성된 골격 포즈의 시퀀스를 취하고, 포즈로 텍스처를 오버레이하여 프레임(122)을 생성한다.

[0024] 도 2는 본 발명의 일 실시예에 따른 인간 행동 생성을 위한 GAN 조직화 방법을 설명하기 위한 흐름도이다.

[0025] 제안하는 인간 행동 생성을 위한 GAN 조직화 방법은 행동 표현부를 통해 골격 포인트의 변화 패턴을 구성하는 인체 움직임을 행동 표현으로 나타내는 단계(210), 포즈 생성부를 통해 행동 표현의 클래스 등급을 입력 받아 행동 표현을 재구성하고 판별하여 포즈를 생성하는 포즈 생성 단계(220), 생성된 포즈의 서로 다른 노이즈 레벨 입력과 예상 스타일 출력을 포즈 최적화부를 통해 쌍으로 매핑하여 노이즈를 제거하고, 순차적인 골격 포즈를 생성하는 포즈 최적화 단계(230) 및 노이즈가 제거된 포즈를 입력 받아 프레임 생성부를 통해 프레임 시퀀스로 변환하는 프레임 생성 단계(240)를 포함한다.

[0026] 단계(210)에서, 행동 표현부를 통해 골격 포인트의 변화 패턴을 구성하는 인체 움직임을 행동 표현으로 나타낸다.

[0027] 적절한 데이터 표현은 특정 업무에서 중요한 역할을 한다. 인간 행동 생성의 경우, 공간에서 골격 포인트의 위치를 학습하는 것만으로 일련의 포즈들을 만들어내려고 하는 것은 효율적이지 않다. 그 대신, 특정한 행동 내에서 골격 포인트의 변화 패턴을 구성하는 인체 움직임의 특별한 표현을 제안한다.

[0028] 도 3은 본 발명의 일 실시예에 따른 행동 표현의 샘플을 나타내는 도면이다.

[0029] 도 3(b)의 포즈 골격이 행동을 취한다고 가정하고(예를 들어, 걷기), 도 3(a)에 해당하는 행동 표현을 도시했다. 도 3(b)의 샘플 포즈는 2D에서 14 포인트 또는 ID 공간에서 28점으로 구성되며, 도 3(a)의 행동 표현은 여러 프레임 내에서 골격 포즈의 각 지점이 어떻게 변화하는지 보여준다.

[0030] 도 3에서는 60프레임 내에서 인간 스트레칭 동작의 예시를 나타내고, $S \times N$ 피처에서 각각의 포인트 움직임을 요약했다. 여기서 N은 프레임의 수이고 S는 원래 포즈 포인트의 변형을 위한 스케일링 계수이다. 이 작업에서는 S를 40으로 적용했다. 머리 움직임의 포인트를 분석하면 x 포인트는 그대로 유지되지만 y는 위치가 바뀐다는 것을 알 수 있다. 따라서 모든 인간의 행동은 제안된 표현으로 전환되어 포즈 생성 단계에서 사용될 것이다.

[0032] 단계(220)에서, 포즈 생성부를 통해 행동 표현의 클래스 등급을 입력 받아 행동 표현을 재구성하고 판별하여 포즈를 생성한다.

[0033] 포즈 생성부의 생성기를 통해 행동 표현의 클래스 등급을 입력 받아 정규적으로 분산된 잠재 차원 노이즈를 획득하며, 임베딩, 클래스 등급에 대한 밀집 계층, 배치 정규화, 잠재 노이즈에 대한 ReLU를 수행한 후 결합을 위해 동일한 차원으로 재구성하여 포즈 생성부의 판별기로 전송한다.

[0034] 이후, 포즈 생성부의 판별기를 통해 생성기와 같은 클래스 등급으로 동일한 단계를 수행하여 생산된 출력과 결합하고, 컨볼루션 블록 계층을 수행하여 판별함으로써 포즈를 생성한다.

[0035] 단계(230)에서, 생성된 포즈의 서로 다른 노이즈 레벨 입력과 예상 스타일 출력을 포즈 최적화부를 통해 쌍으로 매핑하여 노이즈를 제거하고, 순차적인 골격 포즈를 생성한다. 아래에서 포즈 생성 및 최적화에 대해 더욱 상세히 설명한다.

[0036] 먼저, 포즈 생성 및 최적화는 함께 작동하는 두 개의 GAN으로 구성된다. PGAN은 사용자 정의 행동 클래스를 취하여 노이즈로 구현하고 포즈 시퀀스 표현을 생성한다. 다음으로, POGAN은 생성된 시퀀스에 대한 최적화 프로세스를 수행하여 노이즈를 제거하고 PGAN에서 생성된 동작 패턴을 개선한다.

[0037] 본 발명의 실시예에 따른 PGAN 은 GOHAG의 주요 부분으로, 주어진 행동 클래스에 따라 부드러운 포즈 생성을 담당한다. PGAN의 생성기는 행동 표현(도 3(a) 참조)의 형태로 행동을 재현하고 판별기 네트워크를 속이려고 한다. 동시에 판별기는 생성된 포즈가 진짜인지 가짜인지를 식별하는 법을 학습한다.

[0038] 단계(240)에서, 노이즈가 제거된 포즈를 입력 받아 프레임 생성부를 통해 프레임 시퀀스로 변환한다. 골격 이미지를 입력으로 하여 골격의 질감을 오버레이하고, 장면 프레임을 생성한다.

[0040] 도 4는 본 발명의 일 실시예에 따른 인간 행동 생성을 위한 GAN 조직화 장치의 구성을 나타내는 도면이다.

[0041] 제안하는 인간 행동 생성을 위한 GAN 조직화 장치(400)는 행동 표현부(410), 포즈 생성부(420), 포즈 최적화부(430) 및 프레임 생성부(440)를 포함한다.

[0042] 행동 표현부(410)는 골격 포인트의 변화 패턴을 구성하는 인체 움직임을 행동 표현으로 나타낸다.

[0043] 적절한 데이터 표현은 특정 업무에서 중요한 역할을 한다. 인간 행동 생성의 경우, 공간에서 골격 포인트의 위치를 학습하는 것만으로 일련의 포즈들을 만들어내려고 하는 것은 효율적이지 않다. 그 대신, 특정한 행동 내에서 골격 포인트의 변화 패턴을 구성하는 인체 움직임의 특별한 표현을 제안한다

[0044] 포즈 생성부(420)는 행동 표현의 클래스 등급을 입력 받아 행동 표현을 재구성하고 판별하여 포즈를 생성한다. 포즈 생성부(420)는 생성기(421) 및 판별기(424)를 포함한다.

[0045] 포즈 생성부(420)의 생성기(421)를 통해 행동 표현의 클래스 등급을 입력 받아 정규적으로 분산된 잠재 차원 노이즈를 획득하며, 임베딩, 클래스 등급에 대한 밀집 계층, 배치 정규화, 잠재 노이즈에 대한 ReLU를 수행한 후 결합을 위해 동일한 차원으로 재구성하여 포즈 생성부의 판별기로 전송한다.

[0046] 이후, 포즈 생성부(420)의 판별기(422)를 통해 생성기와 같은 클래스 등급으로 동일한 단계를 수행하여 생산된 출력과 결합하고, 컨볼루션 블록 계층을 수행하여 판별함으로써 포즈를 생성한다.

[0047] 포즈 최적화부(430)는 생성된 포즈의 서로 다른 노이즈 레벨 입력과 예상 스타일 출력을 쌍으로 매핑하여 노이즈를 제거하고, 순차적인 골격 포즈를 생성한다.

[0048] 먼저, 포즈 생성 및 최적화는 함께 작동하는 두 개의 GAN으로 구성된다. PGAN은 사용자 정의 행동 클래스를 취하여 노이즈로 구현하고 포즈 시퀀스 표현을 생성한다. 다음으로, POGAN은 생성된 시퀀스에 대한 최적화 프로세스를 수행하여 노이즈를 제거하고 PGAN에서 생성된 동작 패턴을 개선한다.

[0049] 본 발명의 실시예에 따른 PGAN 은 GOHAG의 주요 부분으로, 주어진 행동 클래스에 따라 부드러운 포즈 생성을 담당한다. PGAN의 생성기는 행동 표현(도 3(a) 참조)의 형태로 행동을 재현하고 판별기 네트워크를 속이려고 한다. 동시에 판별기는 생성된 포즈가 진짜인지 가짜인지를 식별하는 법을 학습한다.

[0050] 프레임 생성부(440)는 노이즈가 제거된 포즈를 입력 받아 프레임 시퀀스로 변환한다. 골격 이미지를 입력으로 하여 골격의 질감을 오버레이하고, 장면 프레임을 생성한다.

[0052] 도 5는 본 발명의 일 실시예에 따른 PGAN의 아키텍처를 설명하기 위한 도면이다.

[0053] 제안된 PGAN은 심층 아키텍처를 가지는 조건부 GAN이다. 생성기와 판별기 네트워크는 모두 사용자 정의 클래스 등급으로 조건화되며, 이진 교차 엔트로피 손실에 대해 함께 훈련된다.

[0054] 생성기(510)는 클래스 등급을 받고 정규적으로 분산된 잠재 차원 노이즈(latent dim noise)(본 발명의 실시예를 따르면, 잠재 차원은 100임)를 얻으며, 임베딩, 클래스 등급에 대한 밀집 계층(Dense), 배치 정규화(Batch Norm), 잠재 노이즈에 대한 ReLU를 수행한 후 결합을 위해 동일한 차원으로 재구성(Reshape)한다. 다음으로 컨볼루션2D변환(Conv2DTranspose), 배치 정규화(Batch Norm), ReLU(예를 들어, 블록에서 정의된 컨볼루션용 필터 수, 커널 크기는 5, 스트라이딩(striding)은 2)를 포함하는 일련의 블록을 통과한 출력이다. 그런 다음 네트워

크와 클래스 등급의 출력은 판별기(520) 네트워크로 전달된다.

[0055] 판별기(520) 네트워크는 초기 행동 클래스 등급(522)과 생성기(510)의 생성된 포즈(521)로 작동한다. 그것은 생성기와 같은 행동 등급으로 동일한 단계를 수행하고 생산된 출력과 결합한다. 컨볼루션2D(Conv2D), LeakyReLU, 드롭아웃(Dropout)(본 발명의 실시예를 따르면, LeakyReLU는 0.2, Droput은 0.5)으로 구성된 일련의 컨볼루션 블록 계층을 수행하고, 컨볼루션을 위한 필터의 수를 늘리면(도 5에 도 시함) 판별기가 단일 상태(single statement)(fake 또는 real)를 생성하도록 이끈다. PGAN의 아키텍처는 경험적으로 발전되었다. 본 발명에서는 컨볼루션 계층에서 더 많은 필터들이 항상 더 나은 수렴으로 이어지지는 않을 것이라는 것을 알게 되었고, 생성기와 판별기 네트워크 사이의 균형을 찾는 것이 중요하다. 도 6을 참조 하여 컨볼루션 블록(523)에 대하여 더욱 상세히 설명한다.

[0057] 도 6은 본 발명의 일 실시예에 따른 컨볼루션 블록을 설명하기 위한 도면이다.

[0058] 본 발명의 실시예에 따른 PGAN은 행동의 사용자 등급에서 요청하고 일련의 동작을 생성한다. 정확한 골격 좌표 생성에 초점을 맞추는 대신에, 모델이 특정한 행동의 패턴을 학습하기를 원한다.

[0059] 포즈 생성 네트워크의 주요 아이디어는 데이터 표현이다. 각각의 포즈 지점이 N개의 프레임 수 내에서 어떻게 변화하는지 움직임을 나타낸다. 예를 들어, 30 프레임(N=30)의 기본 표현을 $N \times P$로 하고 여기서 P가 단일 포즈 좌표(P=28)인 경우, 이를 $(N * P) \times D$로 변환하며, D-인덱싱 차원으로 공간에서의 포즈 좌표 위치를 정의한다. 본 발명의 실시예에서는 D를 40으로 사용했다.

[0060] 포즈 생성기의 아키텍처는 도 6에 나타낸 컨볼루션 블록의 스택을 정의한다. 잠재적 차원 노이즈(본 발명의 실시예에 따른 잠재적 차원은 100)를 취하여, 조건부 CGAN(Conditional GAN)이 하듯이 행동 클래스 등급을 내장하고 포즈 생성기의 조건으로 제공한다. 이것이 통과하면서, 네트워크는 다시 $N \times P$ 표현으로 변환하여 $(N * P) \times D$로 결과를 출력한다. 포즈 생성 네트워크의 판별기는 생성기와 구조가 유사하지만 입력으로 $(N * P) \times D$ 형태를 얻고 형태 1(fake 또는 real)로 출력을 반환한다. 생성기는 GAN 에서와 같이 판별기를 속이려 하고, 두 네트워크 모두를 훈련시키기 위해 기본적인 2진 교차 엔트로피 손실을 적용한다.

[0062] 도 7은 본 발명의 일 실시예에 따른 POGAN의 출력을 비교한 결과이다.

[0063] 본 발명의 실시예에 따른 POGAN는 부드럽고 그럴듯한 인간 행동 생성에 상당한 역할을 한다. POGAN의 주요 목적은 PGAN이 생산한 출력을 최적화하는 것이다. 도 7(b)는 PGAN의 샘플 결과를 보여준다. 도 7(a)의 실제 포인트의 패턴과 도 7(b)의 생성된 PGAN을 비교하면 PGAN은 허용 가능한 출력을 생성하지만, 최종 포즈 시퀀스의 부드러움에 영향을 주는 노이즈가 있다. 이는 최적화 단계의 필요성을 입증한다.

[0064] 본 발명의 실시예에 따른 POGAN은 조건부 CycleGAN을 이용한다. CycleGAN은 서로 다른 케이스에서 큰 잠재력과 능력을 보였기 때문에, POGAN에 CycleGAN의 아키텍처를 데이터를 공급하기 위해 채택된 차원으로 적용한다. 이 단계를 스타일 변경 작업으로 간주하여 노이즈가 있는 입력을 매끄러운 포즈로 변환하는 최적화가 주요 목적이다. POGAN을 학습하기 위해 서로 다른 노이즈 레벨 입력과 예상 스타일 출력을 쌍으로 매핑하는 방법을 사용했다. 최적화 단계의 결과, 도 7(b)에 기술된 PGAN의 노이즈된 출력은 도 7(c)에 나타낸 매끄럽고 깨끗한 포즈 표현으로 변환된다. PGAN과 POGAN의 포즈 생성 단계에서의 결과적인 출력은 인간 행동 생성의 다음 단계를 위한 순차적인 골격 포즈를 생성한다.

[0066] 도 8은 본 발명의 일 실시예에 따른 포즈 생성 상태의 샘플 결과를 나타내는 도면이다.

[0067] 본 발명의 실시예에 따른 GOHAG의 프레임 생성 단계에서는 생성된 포즈를 프레임의 시퀀스로 변환한다. FGAN은 주어진 포즈 집합을 위해 각 골격의 질감을 오버레이한다.

[0068] FGAN를 위해 사실적인 이미지 대 이미지 변환을 목적으로 하여NVIDIA의 pix2pixHD를 활용했다. pix2pixHD는 조건부 GAN으로 고해상도 영상 합성을 적용한다. FGAN을 구축하기 위한 기준으로서 pix2pixHD 네트워크를 사용했다.

[0069] pix2pixHD의 입력 계층을 수정하면서, 골격 이미지를 입력으로 하고 장면 프레임을 생성하는 프레임 생성기 GAN을 제안했다. GOHAG의 첫 번째 단계의 출력은 데이터 표현에서 포즈 표현(도 3(a) 참조)이므로, 그것들을 일련의 골격 포즈 이미지로 변환한다(도 8(a) 참조). FGAN에게 골격 포즈 이미지를 제공하면 그 포즈 위의 텍스처

프레임이 돌아온다. FGAN 훈련에는 두 가지 이미지 세트가 필요하다: 첫 번째는 골격 포즈 이미지와 텍스처 프레임 집합이다. FGAN의 성능이 이전 네트워크에서 신뢰할 수 있다는 점에 주목한다. FGAN이 생산하는 결과의 안정성은 다양한 예에 따라 증가한다. 도 8(b)에 도 8(a)의 PGAN&POGAN 생성 포즈를 기반으로 FGAN의 샘플 결과를 나타내었다.

[0070] 포즈 생성에 관한 종래기술과 비교해 보면, GOHAG만의 특징이 있다. 상기 종래기술6에는 네트워크 내부에 LSTM 모듈이 있어 시퀀스 생성을 위해 초기 포즈를 취해야 하지만, GOHAG는 잠재 노이즈만 사용하여 움직임을 생성할 수 있다. 상기 종래기술7에서 연구원들은 포즈 생성 단계에 두 개의 GAN을 적용했다. 한 GAN은 행동의 한 부류를 연결시켜 그것을 포즈로 매핑함으로써 무작위 노이즈를 조작한다.

[0071] 두 번째 GAN은 시퀀스에 따라 노이즈를 발생시킨다. GOHAG의 경우, 움직임의 종류와 결합된 무작위 노이즈를 동작 표현 영상에 매핑하려고 한다(도 3 참조). GOHAG는 특정 움직임을 설명하는 패턴을 가진 이미지 생성으로 포즈 생성을 고려한다.

[0073] 이상에서 설명된 장치는 하드웨어 구성요소, 소프트웨어 구성요소, 및/또는 하드웨어 구성요소 및 소프트웨어 구성요소의 조합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치 및 구성요소는, 예를 들어, 프로세서, 콘트롤러, ALU(arithmetic logic unit), 디지털 신호 프로세서(digital signal processor), 마이크로컴퓨터, FPA(field programmable array), PLU(programmable logic unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 하나 이상의 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다. 처리 장치는 운영 체제(OS) 및 상기 운영 체제 상에서 수행되는 하나 이상의 소프트웨어 애플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응답하여, 데이터를 접근, 저장, 조작, 처리 및 생성할 수도 있다. 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설명된 경우도 있지만, 해당 기술분야에서 통상의 지식을 가진 자는, 처리 장치가 복수 개의 처리 요소(processing element) 및/또는 복수 유형의 처리 요소를 포함할 수 있음을 알 수 있다. 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 콘트롤러를 포함할 수 있다. 또한, 병렬 프로세서(parallel processor)와 같은, 다른 처리 구성(processing configuration)도 가능하다.

[0074] 소프트웨어는 컴퓨터 프로그램(computer program), 코드(code), 명령(instruction), 또는 이들 중 하나 이상의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로(collectively) 처리 장치를 명령할 수 있다. 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성요소(component), 물리적 장치, 가상 장치(virtual equipment), 컴퓨터 저장 매체 또는 장치에 구체화(embody)될 수 있다. 소프트웨어는 네트워크로 연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 하나 이상의 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.

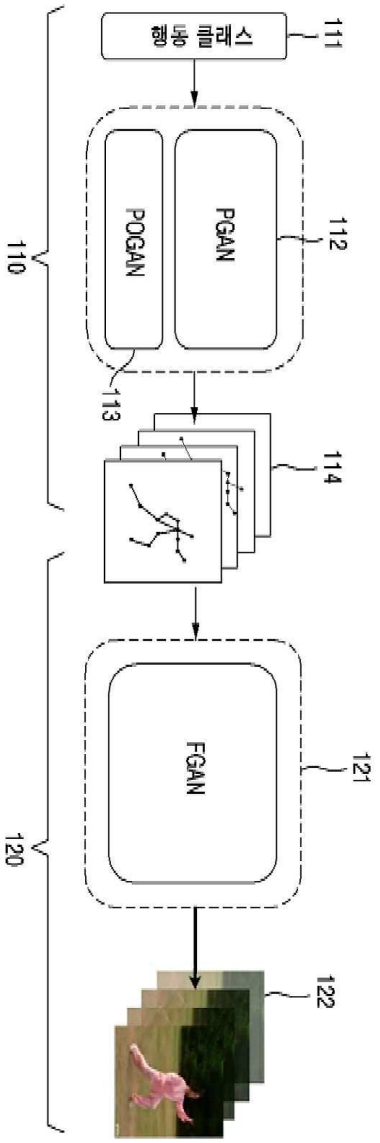
[0075] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다.

[0076] 이상과 같이 실시예들이 비록 한정된 실시예와 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 상기의 기재로부터 다양한 수정 및 변형이 가능하다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다.

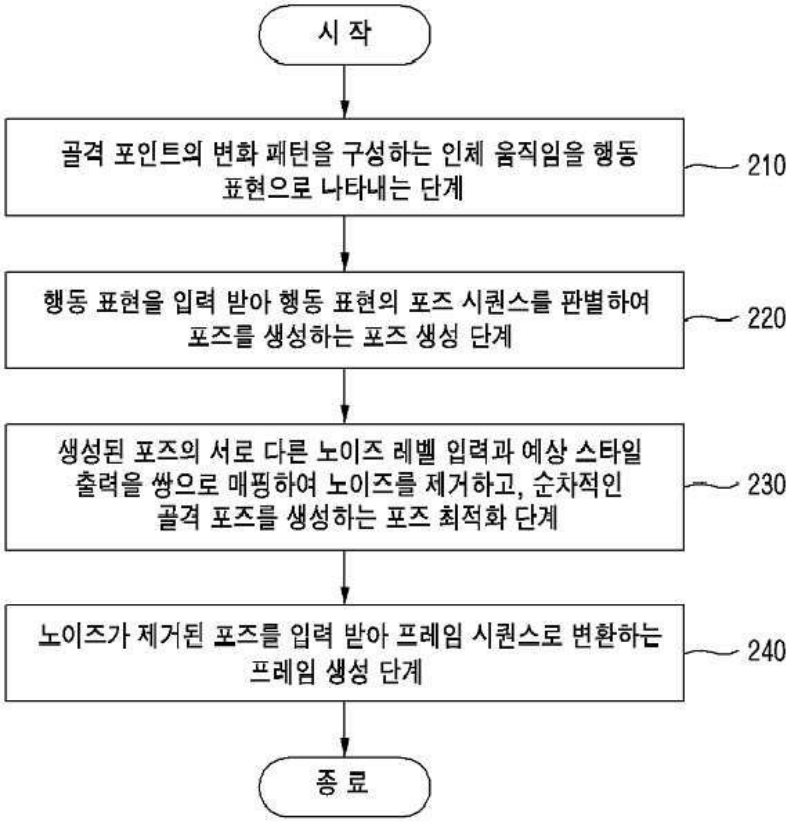
[0077] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

도면

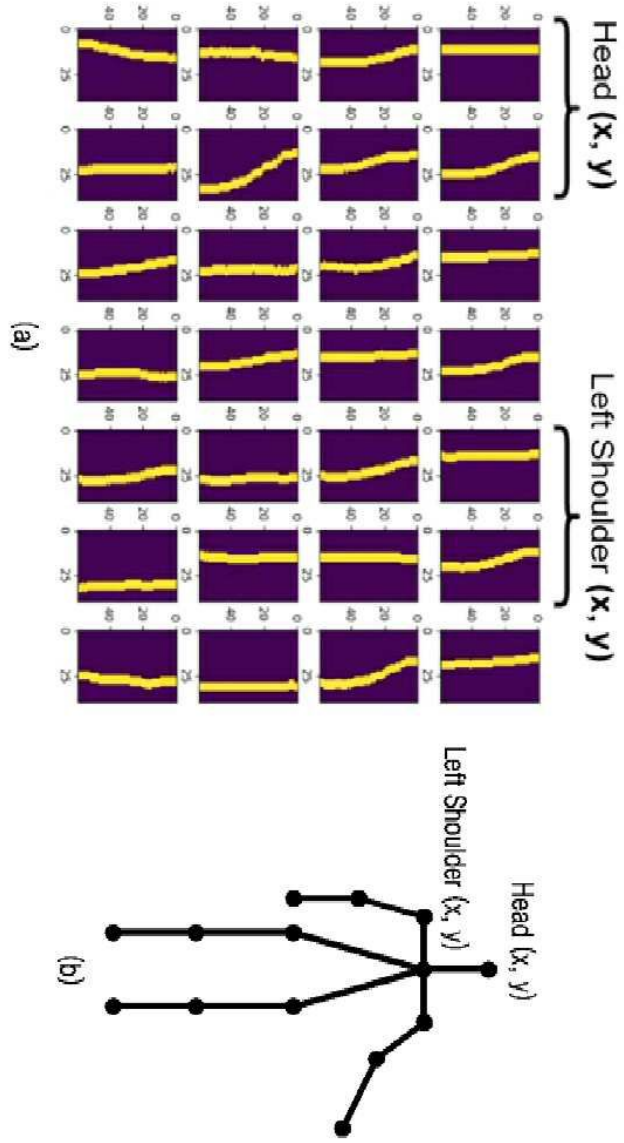
도면1



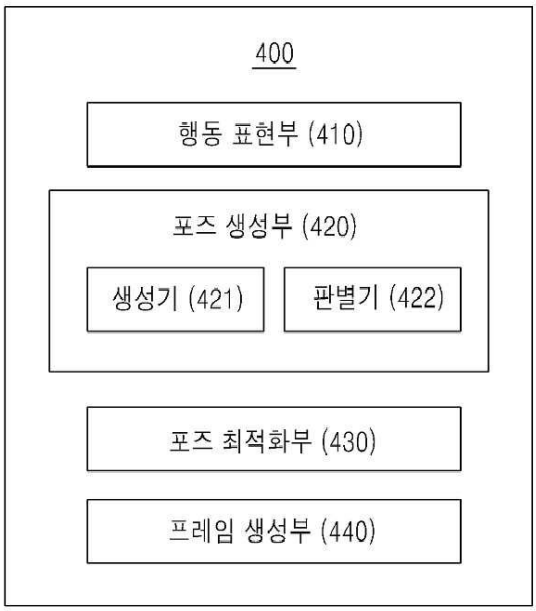
도면2


시 작
골격 포인트의 변화 패턴을 구성하는 인체 움직임을 행동 표현으로 나타내는 단계
210
행동 표현을 입력 받아 행동 표현의 포즈 시퀀스를 판별하여 포즈를 생성하는 포즈 생성 단계
220
생성된 포즈의 서로 다른 노이즈 레벨 입력과 예상 스타일 출력을 쌍으로 매핑하여 노이즈를 제거하고, 순차적인 골격 포즈를 생성하는 포즈 최적화 단계
230
노이즈가 제거된 포즈를 입력 받아 프레임 시퀀스로 변환하는 프레임 생성 단계
240
종 료

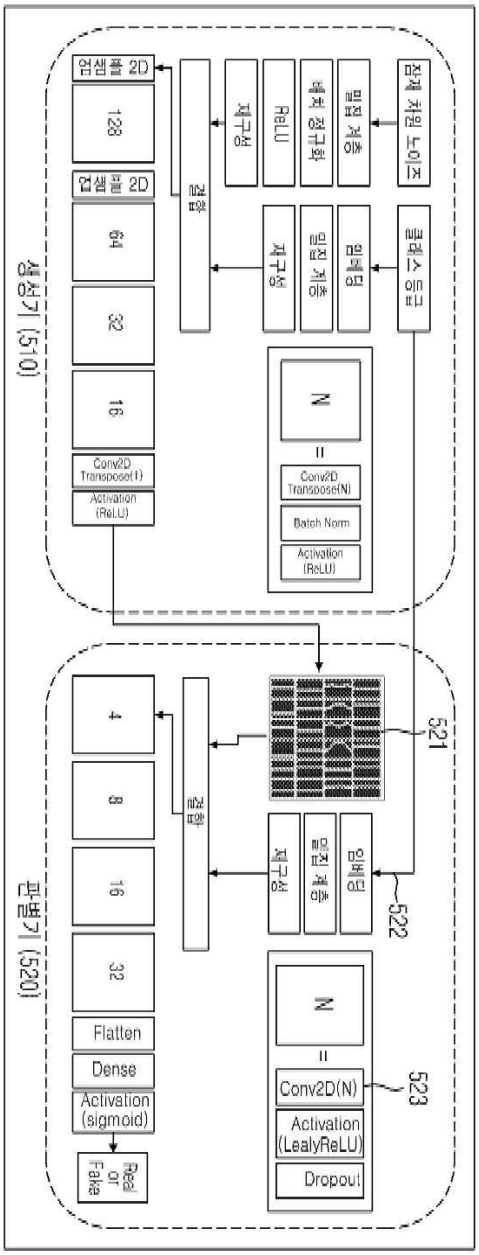
도면3



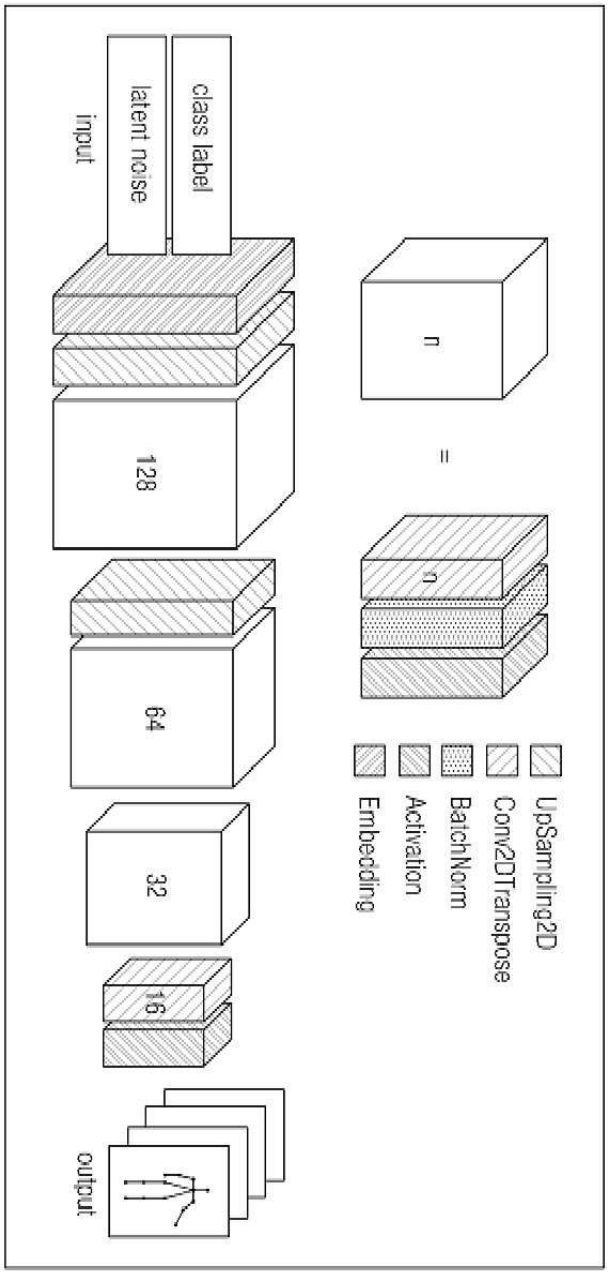
도면4



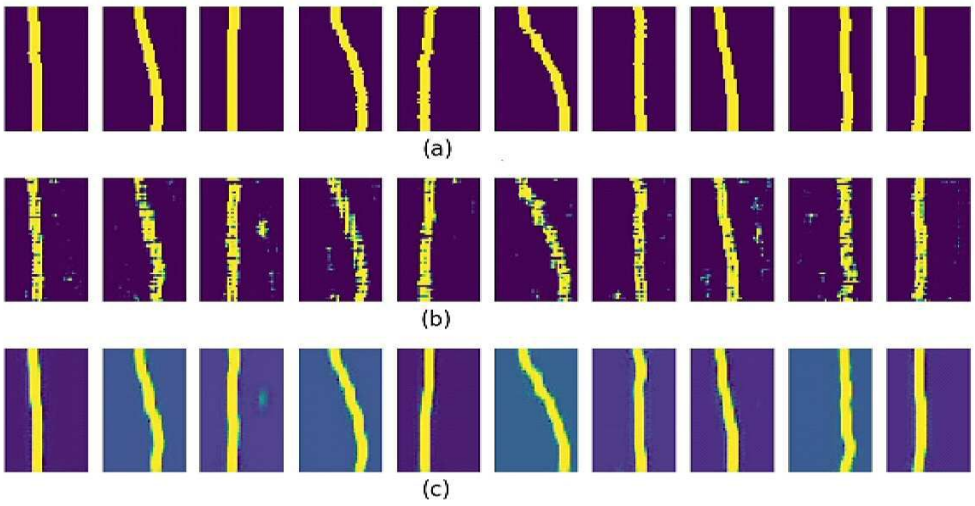
도면5


생성기 (510)
잠재 차원 노이즈
밀집 계층
배치 정규화
ReLU
재구성
결합
클래스 임베딩
밀집 계층
재구성
N
=
Conv2D Transpose(N)
Batch Norm
Activation (ReLU)
Conv2D Transpose(1)
Activation (ReLU)
128
64
32
16
판별기 (520)
521
522
523
임베딩
밀집 계층
재구성
결합
N
=
Conv2D(N)
Activation (LealyReLU)
Dropout
4
8
16
32
Flatten
Dense
Activation (sigmoid)
Real or Fake

도면6


input
latent noise
class label
128
64
32
16
output
n
=
n
Embedding
Activation
BatchNorm
Conv2DTranspose
UpSampling2D

도면7



도면8

