



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2021년04월06일

(11) 등록번호 10-2237124

(24) 등록일자 2021년04월01일

(51) 국제특허분류(Int. Cl.)

G06T 13/40 (2011.01) G06N 3/04 (2006.01)

G06T 7/20 (2017.01)

(52) CPC특허분류

G06T 13/40 (2013.01)

G06N 3/0454 (2013.01)

(21) 출원번호 10-2020-0018278

(22) 출원일자 2020년02월14일

심사청구일자 2020년02월14일

(56) 선행기술조사문헌

KR1020100002803 A*

Cheema et al. Fine-Grained Semantic Segmentation of Motion Capture Data using Dilated Temporal Fully-Convolutional Networks. arXiv preprint, 2019년 3월, pp. 1-4. 1부.*

Jang et al. A variational U-Net for motion retargeting. SIGGRAPH Asia, 2018년 12월, pp. 1-9. 1부.*

Zhu et al. Unpaired image-to-image translation using cycle-consistent adversarial networks. ICCV, 2017년 10월, pp. 2223-2232. 1부.*

*는 심사관에 의하여 인용된 문헌

(73) 특허권자

고려대학교 산학협력단

서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)

(72) 발명자

한정현

서울특별시 성북구 종암로23길 35, 210동 301호(종암동, 래미안세레니티)

김상빈

서울특별시 동대문구 안암로18길 16, 501호(제기동)

(뒷면에 계속)

(74) 대리인

송인호, 윤형근, 최관락

전체 청구항 수 : 총 4 항

심사관 : 김창원

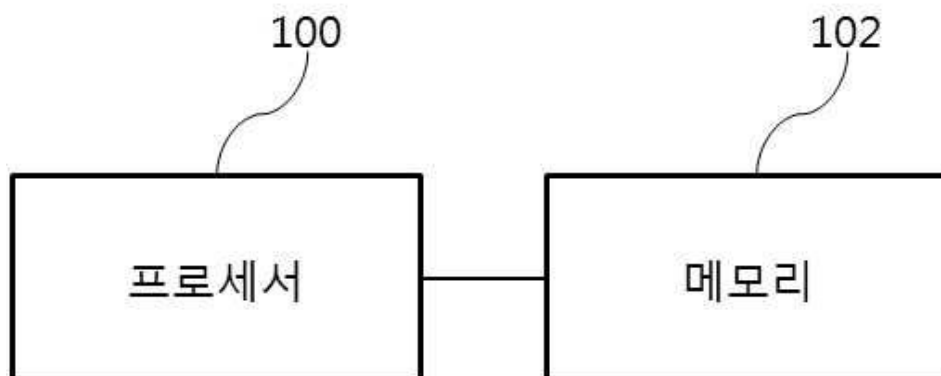
(54) 발명의 명칭 신경망 기반 모션 리타겟팅 방법 및 장치

(57) 요약

본 발명은 신경망 기반 모션 리타겟팅 방법 및 장치를 개시한다. 본 발명에 따르면, 프로세서; 및 상기 프로세서에 연결되는 메모리를 포함하되, 상기 메모리는, 미리 설정된 개수의 연속되는 프레임에서의 소스 캐릭터의 소스 모션 시퀀스, 상기 소스 캐릭터의 소스 뼈대(skeleton)에 관한 정보 및 리타겟팅 대상이 되는 타겟 캐릭터의 타

(뒷면에 계속)

대표도 - 도1



겟 뼈대에 관한 정보를 입력 받고-상기 소스 모션 시퀀스는, 상기 소스 뼈대의 관절 회전 정보 및 상기 소스 뼈대의 루트 관절 회전 정보를 포함함-, 정기구학(forward kinematics, FK) 계산을 통해 상기 소스 뼈대에 관한 정보 및 상기 소스 뼈대의 관절 회전 정보를 이용하여 상기 소스 뼈대의 관절 로컬 좌표를 획득하고, 상기 소스 뼈대의 관절 회전 정보, 루트 관절 회전 정보, 관절 로컬 좌표 및 상기 타겟 뼈대에 관한 정보를 합성곱 신경망 모델에 입력하여 상기 타겟 캐릭터의 타겟 모션 시퀀스를 출력하도록, 상기 프로세서에 의해 실행 가능한 프로그램 명령어들을 저장하는 신경망 기반 모션 리타겟팅 장치가 제공된다.

(52) CPC특허분류

G06T 7/20 (2013.01)
G06T 2207/20044 (2013.01)
G06T 2207/20081 (2013.01)
G06T 2207/20084 (2013.01)

권성수

서울특별시 동대문구 제기로 40-10, 105호(제기동)

(72) 발명자

박인범

서울특별시 마포구 연남로 27, 201호(연남동, 청기
 와아파트)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711097540
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	차세대정보·컴퓨팅기술개발(R&D)
연구과제명	초실감 원격가상 인터랙션을 위한 사용자 복원 기술 개발
기 여 율	1/1
과제수행기관명	고려대학교 산학협력단
연구기간	2019.06.01 ~ 2020.03.31

명세서

청구범위

청구항 1

신경망 기반 모션 리타겟팅 장치로서,

프로세서; 및

상기 프로세서에 연결되는 메모리를 포함하되,

상기 메모리는,

미리 설정된 개수의 연속되는 프레임에서의 소스 캐릭터의 소스 모션 시퀀스, 상기 소스 캐릭터의 소스 뼈대(skeleton)에 관한 정보 및 리타겟팅 대상이 되는 타겟 캐릭터의 타겟 뼈대에 관한 정보를 입력 받고-상기 소스 모션 시퀀스는, 상기 소스 뼈대의 관절 회전 정보 및 상기 소스 뼈대의 루트 관절 회전 정보를 포함함-,

정기구학(forward kinematics, FK) 계산을 통해 상기 소스 뼈대에 관한 정보 및 상기 소스 뼈대의 관절 회전 정보를 이용하여 상기 소스 뼈대의 관절 로컬 좌표를 획득하고,

상기 소스 뼈대의 관절 회전 정보, 루트 관절 회전 정보, 관절 로컬 좌표 및 상기 타겟 뼈대에 관한 정보를 합성곱 신경망 모델에 입력하여 상기 타겟 캐릭터의 타겟 모션 시퀀스를 출력하도록,

상기 프로세서에 의해 실행 가능한 프로그램 명령어들을 저장하되,

상기 합성곱 신경망 모델은 TDCN(Temporal dilated convolutional network)이고, 상기 TDCN은 패치 기반 GAN(Generative Adversarial Network)으로 사이클 일관성이 훈련되고,

상기 훈련의 목적함수는 사이클 일관성 손실, 관절 비트맵 손실, 높이(height) 손실, 적대적 손실, 정규화 손실 및 방향 손실 중 적어도 하나를 포함하는 신경망 기반 모션 리타겟팅 장치.

청구항 2

삭제

청구항 3

제1항에 있어서,

상기 TDCN은 상기 연속되는 프레임에서 상기 타겟 뼈대의 관절 회전 정보 및 상기 타겟 뼈대의 루트 관절 회전 정보를 출력하며,

상기 프로그램 명령어들은,

정기구학(forward kinematics, FK) 계산을 통해 상기 타겟 뼈대 정보에 관한 정보 및 상기 타겟 뼈대의 관절 회전 정보를 이용하여 상기 타겟 뼈대의 관절 로컬 좌표를 획득하고,

상기 타겟 뼈대의 관절 로컬 좌표 및 상기 타겟 뼈대의 루트 관절 회전 정보를 이용하여 상기 타겟 모션 시퀀스를 생성하는 신경망 기반 모션 리타겟팅 장치.

청구항 4

삭제

청구항 5

삭제

청구항 6

제1항에 있어서,

상기 높이 손실은, 상기 소스 캐릭터와 상기 타겟 캐릭터의 높이 차이에 의한 손실을 반영하며,

상기 높이 손실은, 서로 인접한 프레임 사이에서 정의되는 상기 소스 캐릭터의 로컬 모션의 차이(제1 차이)와 상기 타겟 캐릭터의 관절 로컬 좌표 차이(제2 차이) 각각을 상기 소스 캐릭터의 높이와 상기 타겟 캐릭터의 높이를 이용하여 정규화하여 계산된 값을 이용하여 결정되는 신경망 기반 모션 리타겟팅 장치.

청구항 7

프로세서 및 메모리를 포함하는 장치에서 신경망 기반으로 모션 리타겟팅하는 방법으로서,

미리 설정된 개수의 연속되는 프레임에서의 소스 캐릭터의 소스 모션 시퀀스, 상기 소스 캐릭터의 소스 뼈대(skeleton)에 관한 정보 및 리타겟팅 대상이 되는 타겟 캐릭터의 타겟 뼈대에 관한 정보를 입력 받는 단계-상기 소스 모션 시퀀스는, 상기 소스 뼈대의 관절 회전 정보 및 상기 소스 뼈대의 루트 관절 회전 정보를 포함함-;

정기구학(forward kinematics, FK) 계산을 통해 상기 소스 뼈대에 관한 정보 및 상기 소스 뼈대의 관절 회전 정보를 이용하여 상기 소스 뼈대의 관절 로컬 좌표를 획득하는 단계; 및

상기 소스 뼈대의 관절 회전 정보, 루트 관절 회전 정보, 관절 로컬 좌표 및 상기 타겟 뼈대에 관한 정보를 합성곱 신경망 모델에 입력하여 상기 타겟 캐릭터의 타겟 모션 시퀀스를 출력하는 단계를 포함하되,

상기 합성곱 신경망 모델은 TDCN(Temporal dilated convolutional network)이고, 상기 TDCN은 패치 기반 GAN(Generative Adversarial Network)으로 사이클 일관성이 훈련되고,

상기 훈련의 목적함수는 사이클 일관성 손실, 관절 비틀림 손실, 높이(height) 손실, 적대적 손실, 정규화 손실 및 방향 손실 중 적어도 하나를 포함하는 신경망 기반 모션 리타겟팅 방법.

발명의 설명

기술 분야

[0001] 본 발명은 신경망 기반 모션 리타겟팅 방법 및 장치에 관한 것으로서, 보다 상세하게는, 딥러닝을 이용하여 캐릭터 애니메이션 생성하는 방법 및 장치에 관한 것이다.

배경 기술

[0002] 모션 리타겟팅은 원본 캐릭터(소스)의 모션을 뼈대(skeleton)의 크기가 다른 타겟 캐릭터에 맞추어 주는 과정이다. 리타겟팅된 모션은 자연스러워야 하며 소스 모션의 특징을 보존해야만 한다.

[0003] Villegas 등은 서로 다른 뼈대(skeleton)를 갖는 캐릭터 사이에 자연스러운 모션 리타겟팅을 수행하기 위한 알고리즘을 제안하였다.

[0004] 제안된 알고리즘은 리타겟팅을 수행할 새로운 타겟의 뼈대(target skeleton)와 소스 모션의 각 관절에 대한 위치 정보를 RNN(Recurrent Neural Network)에 입력하여 리타겟팅된 모션을 출력한다.

[0005] 딥러닝 모델을 지도 학습(supervised learning)하기 위해서는 소스 모션과 리타겟팅된 모션의 쌍으로 이루어진 데이터를 다중-다량으로 확보해야 하나, 이를 위해서는 매우 많은 시간과 노력이 필요하다.

[0006] 종래기술의 RNN 기반 모션 리타겟팅은 학습 과정이 불안정하고, 학습 속도를 최적화하기 어려우며, 학습을 진행할 때, 뼈대의 길이를 명시적으로 고려하지 못하여 서로 다른 크기를 가진 캐릭터들의 사이에 리타겟팅 결과가 자연스럽게 않은 문제점이 있다.

[0007] 또한, 종래에는 캐릭터의 루트 관절(hip point)의 위치 차이(position offset)를 출력하기 때문에 오차가 매 프레임 발생하여 누적되므로 짧은 모션이 아닌 긴 모션에 대해 리타겟팅을 수행하면 매우 큰 오차가 발생하는 문제점이 있다.

선행기술문헌

비특허문헌

- [0008] (비특허문헌 0001) Villegas, Ruben, et al. "Neural kinematic networks for unsupervised motion retargetting." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018. DOI:10.1109/CVPR.2018.00901.

발명의 내용

해결하려는 과제

- [0009] 상기한 종래기술의 문제점을 해결하기 위해, 본 발명은 안정적이고 빠른 학습을 수행하고, 서로 다른 크기의 캐릭터들 사이에 자연스러운 리타겟팅을 실현하며, 리타겟팅 대상 모션 길이에 제한이 없는 신경망 기반 모션 리타겟팅 방법 및 장치를 제안하고자 한다.

과제의 해결 수단

- [0010] 상기한 바와 같은 목적을 달성하기 위하여, 본 발명의 일 실시예에 따르면, 신경망 기반 모션 리타겟팅 장치로서, 프로세서; 및 상기 프로세서에 연결되는 메모리를 포함하되, 상기 메모리는, 미리 설정된 개수의 연속되는 프레임에서의 소스 캐릭터의 소스 모션 시퀀스, 상기 소스 캐릭터의 소스 뼈대(skeleton)에 관한 정보 및 리타겟팅 대상이 되는 타겟 캐릭터의 타겟 뼈대에 관한 정보를 입력 받고-상기 소스 모션 시퀀스는, 상기 소스 뼈대의 관절 회전 정보 및 상기 소스 뼈대의 루트 관절 회전 정보를 포함함-, 정기구학(forward kinematics, FK) 계산을 통해 상기 소스 뼈대에 관한 정보 및 상기 소스 뼈대의 관절 회전 정보를 이용하여 상기 소스 뼈대의 관절 로컬 좌표를 획득하고, 상기 소스 뼈대의 관절 회전 정보, 루트 관절 회전 정보, 관절 로컬 좌표 및 상기 타겟 뼈대에 관한 정보를 합성곱 신경망 모델에 입력하여 상기 타겟 캐릭터의 타겟 모션 시퀀스를 출력하도록, 상기 프로세서에 의해 실행 가능한 프로그램 명령어들을 저장하는 신경망 기반 모션 리타겟팅 장치가 제공된다.
- [0011] 상기 합성곱 신경망 모델은 TDCN(Temporal dilated convolutional network)일 수 있다.
- [0012] 상기 TDCN은 상기 연속되는 프레임에서 상기 타겟 뼈대의 관절 회전 정보 및 상기 타겟 뼈대의 루트 관절 회전 정보를 출력하며, 상기 프로그램 명령어들은, 정기구학(forward kinematics, FK) 계산을 통해 상기 타겟 뼈대 정보에 관한 정보 및 상기 타겟 뼈대의 관절 회전 정보를 이용하여 상기 타겟 뼈대의 관절 로컬 좌표를 획득하고, 상기 타겟 뼈대의 관절 로컬 좌표 및 상기 타겟 뼈대의 루트 관절 회전 정보를 이용하여 상기 타겟 모션 시퀀스를 생성할 수 있다.
- [0013] 상기 TDCN은 패치 기반 GAN(Generative Adversarial Network)으로 사이클 일관성이 훈련될 수 있다.
- [0014] 상기 훈련의 목적함수는 사이클 일관성 손실, 관절 비트맵 손실, 높이(height) 손실, 적대적 손실, 정규화 손실 및 방향 손실 중 적어도 하나를 포함할 수 있다.
- [0015] 상기 높이 손실은, 상기 소스 캐릭터와 상기 타겟 캐릭터의 높이 차이에 의한 손실을 반영하며, 상기 높이 손실은, 서로 인접한 프레임 사이에서 정의되는 상기 소스 캐릭터의 로컬 모션의 차이(제1 차이)와 상기 타겟 캐릭터의 로컬 모션의 차이(제2 차이) 각각을 상기 소스 캐릭터의 높이와 상기 타겟 캐릭터의 높이를 이용하여 정규화하여 계산된 값을 이용하여 결정될 수 있다.
- [0016] 본 발명의 다른 측면에 따르면, 프로세서 및 메모리를 포함하는 장치에서 신경망 기반으로 모션 리타겟팅하는 방법으로서,
- [0017] 미리 설정된 개수의 연속되는 프레임에서의 소스 캐릭터의 소스 모션 시퀀스, 상기 소스 캐릭터의 소스 뼈대(skeleton)에 관한 정보 및 리타겟팅 대상이 되는 타겟 캐릭터의 타겟 뼈대에 관한 정보를 입력 받는 단계-상기 소스 모션 시퀀스는, 상기 소스 뼈대의 관절 회전 정보 및 상기 소스 뼈대의 루트 관절 회전 정보를 포함함-; 정기구학(forward kinematics, FK) 계산을 통해 상기 소스 뼈대에 관한 정보 및 상기 소스 뼈대의 관절 회전 정보를 이용하여 상기 소스 뼈대의 관절 로컬 좌표를 획득하는 단계; 및 상기 소스 뼈대의 관절 회전 정보, 루트 관절 회전 정보, 관절 로컬 좌표 및 상기 타겟 뼈대에 관한 정보를 합성곱 신경망 모델에 입력하여 상기 타겟 캐릭터의 타겟 모션 시퀀스를 출력하는 단계를 포함하는 신경망 기반 모션 리타겟팅 방법이 제공된다.

발명의 효과

- [0018] 본 발명에 따르면 Temporal Dilated Convolutional Network (TDCN)을 사용하여 기존 Convolutional Neural

Network (CNN) 대비 적은 파라미터를 가짐에도 불구하고, 높은 성능을 보이면서 학습 과정이 안정적이고 빠른 장점이 있다.

[0019] 또한, 본 발명에 따르면 학습 과정에서 서로 다른 크기의 캐릭터 뼈대를 명시적으로 다룰 수 있도록 캐릭터 뼈대 기반(skeleton-specific) 손실 함수를 고안하여 서로 다른 크기의 캐릭터들 사이에 자연스러운 리타겟팅을 수행할 수 있다.

[0020] 나아가, 기존 캐릭터 애니메이션에서 사용되는 데이터를 그대로 사용하여 입력 데이터에 대한 전처리 과정을 없앨 수 있고, 루트 관절의 위치를 직접 출력(regression)하여 모션 길이에 제한이 없도록 개선하면서도 PatchGAN을 도입하여 리타겟팅 결과물의 디테일이 우수한 장점이 있다.

도면의 간단한 설명

[0021] 도 1은 본 발명의 일 실시예에 따른 신경망 기반 모션 리타겟팅 장치를 도시한 도면이다.

도 2는 본 발명의 일 실시예에 따른 전체 아키텍처를 도시한 도면이다.

도 3은 본 실시예에 따른 TDCN의 구조를 도시한 도면이다.

도 4는 본 실시예에 따른 판별기의 구조를 도시한 도면이다.

도 5는 본 실시예에 따른 사이클을 나타낸 도면이다.

발명을 실시하기 위한 구체적인 내용

[0022] 본 발명은 다양한 변경을 가할 수 있고 여러 가지 실시예를 가질 수 있는 바, 특정 실시예들을 도면에 예시하고 상세하게 설명하고자 한다.

[0023] 그러나, 이는 본 발명을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다.

[0025] 본 발명은 서로 다른 크기(height, 높이)의 캐릭터들 사이에 자연스러운 리타겟팅을 위한 방법을 제안한다.

[0026] 도 1은 본 발명의 일 실시예에 따른 신경망 기반 모션 리타겟팅 장치를 도시한 도면이다.

[0027] 도 1에 도시된 바와 같이, 본 실시예에 따른 신경망 기반 모션 리타겟팅 장치는 프로세서(100) 및 메모리(102)를 포함할 수 있다.

[0028] 프로세서(100)는 컴퓨터 프로그램을 실행할 수 있는 CPU(central processing unit)나 그밖에 가상 머신 등을 포함할 수 있다.

[0029] 메모리(102)는 고정식 하드 드라이브나 착탈식 저장 장치와 같은 불휘발성 저장 장치를 포함할 수 있다. 착탈식 저장 장치는 콤팩트 플래시 유닛, USB 메모리 스틱 등을 포함할 수 있다. 메모리(102)는 각종 랜덤 액세스 메모리와 같은 휘발성 메모리도 포함할 수 있다.

[0030] 본 발명의 일 실시예에 따르면, 메모리(102)에는 소스 캐릭터의 모션 시퀀스 및 타겟 캐릭터의 뼈대 정보를 합성곱 신경망 모델에 입력하여 타겟 캐릭터의 타겟 모션 시퀀스를 출력하는 프로그램 명령어들이 저장된다.

[0031] 보다 상세하게, 본 실시예에 따른 프로그램 명령어들은, 미리 설정된 개수의 연속되는 프레임에서의 소스 캐릭터의 소스 모션 시퀀스, 상기 소스 캐릭터의 소스 뼈대(skeleton)에 관한 정보 및 리타겟팅 대상이 되는 타겟 캐릭터의 타겟 뼈대에 관한 정보를 입력 받는다.

[0032] 여기서, 상기 소스 모션 시퀀스는, 상기 소스 뼈대의 관절 회전 정보 및 상기 소스 뼈대의 루트 관절 회전 정보를 포함할 수 있다.

[0033] 또한, 프로그램 명령어들은, 정기구학(forward kinematics, FK) 계산을 통해 상기 소스 뼈대에 관한 정보 및 상기 소스 뼈대의 관절 회전 정보를 이용하여 상기 소스 뼈대의 관절 로컬 좌표를 획득하고, 상기 소스 뼈대의 관절 회전 정보, 루트 관절 회전 정보, 관절 로컬 좌표 및 상기 타겟 뼈대에 관한 정보를 합성곱 신경망 모델에 입력하여 상기 타겟 캐릭터의 타겟 모션 시퀀스를 출력한다.

[0034] 여기서, 합성곱 신경망 모델은 TDCN(Temporal dilated convolutional network)일 수 있다. 이에 대해서는 하기에서 상세하게 설명될 것이다.

- [0035] 본 실시예에 따른 TDCN은 패치 기반 GAN(Generative Adversarial Network)으로 사이클 일관성이 훈련된다. 여기서, 훈련의 목적함수는 사이클 일관성 손실, 관절 비틀림 손실, 높이(height) 손실, 적대적 손실, 정규화 손실 및 방향 손실 중 적어도 하나를 포함할 수 있다.
- [0036] 서로 다른 크기(높이)를 갖는 소스와 타겟 캐릭터 사이의 자연스러운 리타겟팅을 위해 높이 손실(height loss)이 고려된다.
- [0037] 이하에서는 도면을 참조하여 본 실시예에 따른 리타겟팅을 상세하게 설명한다.
- [0038] 도 2는 본 발명의 일 실시예에 따른 전체 아키텍처를 도시한 도면이다.
- [0039] 도 2를 참조하면, 소스 캐릭터에 의해 수행되는 모션 시퀀스가 TDCN(Temporal dilated convolutional network)을 통해 타겟 캐릭터에 매핑되는 과정을 도시한 도면이다.
- [0040] 도 2에서 소스 캐릭터와 타겟 캐릭터는 각각 A 및 B로 표기하며, 이들은 동일한 수의 관절(joint)을 가진다.

[0041] TDCN의 입력은 소스 캐릭터의 소스 모션 시퀀스 $x_{1:T}^A$ (여기서, T는 프레임 수), 소스 캐릭터의 뼈대에 관한 정보 $\bar{s}^A$ 와 디폴트 포즈의 타겟 캐릭터의 타겟 뼈대에 관한 정보 $\bar{s}^B$ 로 구성된다.

[0042] 소스 모션 시퀀스는 [1, T]의 각 프레임 t에서 소스 모션(source motion) x_t^A 으로 구성되고, 각 소스 모션 x_t^A 는 q_t^A 및 r_t^A 의 조합으로 구성된다.

[0043] 여기서, $q_t^A (\in \mathbb{R}^{4N} \text{ for } N \text{ joints})$ 는 소스 뼈대의 관절 회전 정보를 나타내는 단위 쿼터니언이며, $r_t^A (\in \mathbb{R}^3)$ 는 소스 뼈대의 루트 관절의 위치를 나타낸다.

[0044] q_t^A 및 $\bar{s}^A$ (디폴트 포즈의 소스 뼈대)를 입력으로 하여 정기구학(forward kinematics, FK) 계산을 수행하여 소스 뼈대의 관절 로컬 좌표 $p_t^A (\in \mathbb{R}^{3N})$ 를 얻는다.

[0045] 이는 아래의 수학적식으로 표현된다.

수학적식 1

$$p_t^A = f_{FK}(q_t^A, \bar{s}^A)$$

[0047] 이후 알고리즘에서는 전체 모션 프레임을 사용한다.

[0048] 본 실시예에 따르면, TDCN(Temporal dilated convolutional network)에 대한 입력은 프레임의 모션 시퀀스이고, 각 프레임은 소스의 p_t^A , q_t^A 및 r_t^A 와 타겟의 $\bar{s}^B$ 를 포함한다.

[0049] 매 입력 프레임에서 리타겟팅된 모션을 생성하기 위해, 입력 시퀀스의 왼쪽 끝은 p_1^A , q_1^A 및 r_1^A 이, 오른쪽 끝은 p_t^A , q_t^A 및 r_t^A 가 채워진다.

[0050] $\tilde{x}_{1:T}^A$ 로 표시된 소스 모션 시퀀스, 그리고 타겟 뼈대 $\tilde{s}^B$ 는 TDCN에 입력되고, 타겟 캐릭터의 타겟 모션 시퀀스 $\hat{x}_{1:T}^B$ 가 출력된다.

수학식 2

[0051]
$$\hat{x}_{1:T}^B = f_{TDCN}(\tilde{x}_{1:T}^A, \tilde{s}^B)$$

[0052] 여기서, 각 성분 $\hat{x}_t^B$ 는 타겟 뼈대의 루트 관절의 글로벌 위치 $\hat{r}_t^B (\in \mathbb{R}^3)$ 와 타겟 뼈대의 관절 회전 정보인 단위 쿼터니언 $\hat{q}_t^B (\in \mathbb{R}^{4N})$ 의 조합이다. FK 모듈은 $\hat{q}_t^B$ 를 $\hat{p}_t^B$ 로 변환하고, 이는 타겟 뼈대의 관절 로컬 좌표를 나타낸다.

수학식 3

[0053]
$$\hat{p}_{1:T}^B = f_{FK}(\hat{q}_{1:T}^B, \tilde{s}^B)$$

[0054] 타겟 모션에서 관절의 글로벌 좌표 결정을 위해 $\hat{r}_t^B$ 와 결합된다.

[0055] 도 3은 본 실시예에 따른 TDCN의 구조를 도시한 도면이다.

[0056] 본 실시예에 따른 TDCN은 입력 프레임의 p_t^A , q_t^A , r_t^A 및 $\tilde{s}^B$ 에 컨볼루션 레이어를 적용한다.

[0057] 도 3에서, 제1 컨볼루션 레이어는 "3d1, 1024"로 표시된다. 제1 요소는 필터 크기가 3, 즉, 3개의 프레임이 얹혀있고, 팽창 인자(dilation factor)는 1임을 의미한다. 두 번째 요소는 1024개의 출력 채널을 나타낸다.

[0058] 컨볼루션 레이어 다음에는 3개의 ResNet 타입 블록이 있다. 각 블록은 잔차를 대칭적으로 슬라이스 하고, 후속 특징에 추가하는 스킵 연결로 둘러싸여 있다.

[0059] i번째 블록에서, 컨볼루션의 팽창 인자는 3^i 이다. 각 블록에서, 이러한 컨볼루션 다음에 1d1으로 표시된 선형 투영이 이어진다.

[0060] TDCN에서 수용 필드는 81로 제한되고, 프레임, 즉 TDCN은 한 번에 81개의 프레임의 로컬 시퀀스를 볼 수 있다. 개념적으로 보면 윈도우는 $\tilde{x}_{1:T}^A$ 위로 슬라이드 되고, $\hat{x}_t^B$ 는 상기한 윈도우를 통해 볼 수 있는 소스 모션 시퀀스를 이용하여 생성된다.

[0061] 도 3의 7개의 레이어 각각 다음에는 배치 정규화(batch normalization), LReLU(leaky rectified linear unit) 및 드롭아웃(dropout)이 이어진다. 도 2에서는 간략화를 위해 이들을 도시하지 않았다.

[0062] 마지막 블록에 의해 생성된 특징은 추가 컨볼루션 레이어에 의해 상위 레벨 형상으로 변환된다. 마지막으로, 풀리-커넥티드 레이어(fully-connected layer)는 상위 레벨 특징을 타겟의 $\hat{q}_t^B$ 및 $\hat{r}_t^B$ 에 매핑한다.

[0063] 본 실시예에 따르면, TDCN은 패치 기반 GAN(Generative Adversarial Network)으로 사이클 일관성이 훈련된다.

[0064] 적대적 학습을 위해, 도 3의 TDCN은 생성기(Generator)로 동작한다.

[0065] 판별기(Discriminator)를 위한 리타겟팅된 모션 시퀀스가 제공되고, 이는 시간적으로 확장된 컨볼루션 네트워크이다.

[0066] 도 4는 본 실시예에 따른 판별기의 구조를 도시한 도면이다.

[0067] 도 4의 판별기는 도 3의 생성기와 유사하나 수용 필드는 31 프레임으로 제한된다.

[0068] 종래의 Villegas 등이 제안한 프레임워크에서 전체 모션 시퀀스가 판별기로 입력되는데, 이런 방식으로는 로컬 모션의 고주파 디테일이 제대로 생성되지 않을 수 있다.

[0069] 이하에서는 본 실시예에 따른 생성기와 판별기를 상세하게 설명한다.

[0070] 생성기(G)는 사이클 일관성(cycle consistency) 훈련을 위해 A의 모션 시퀀스 $x_{1:T}^A$ 를 B로 리타겟팅하여 $\hat{x}_{1:T}^B$ 를 생성한 다음, $\hat{x}_{1:T}^B$ 를 A로 리타겟팅하여 $\hat{x}_{1:T}^A$ 를 생성한다.

수학식 4

$$[0071] \quad \hat{x}_{1:T}^B = G(x_{1:T}^A, s^B)$$

수학식 5

$$[0072] \quad \hat{x}_{1:T}^A = G(\hat{x}_{1:T}^B, s^A)$$

[0073] 도 5는 본 실시예에 따른 사이클을 나타낸 도면이다.

[0074] 본 실시예에서는 6개의 손실 항을 이용한다. $\mathcal{L}_c$ 는 사이클 일관성 손실, $\mathcal{L}_t$ 는 관절 비틀림 손실, $\mathcal{L}_h$ 는 높이 손실, $\mathcal{L}_a$ 는 적대적 손실, $\mathcal{L}_r$ 는 정규화 손실, $\mathcal{L}_o$ 는 방향(orientation) 손실이다.

[0075] 전체 훈련 목적 함수는 다음과 같이 정의된다.

수학식 6

$$[0076] \quad \min_G \max_D \mathcal{L}_c + \lambda_t \mathcal{L}_t + \lambda_h \mathcal{L}_h + \lambda_a \mathcal{L}_a + \lambda_r \mathcal{L}_r + \lambda_o \mathcal{L}_o$$

[0077] 여기서, λ_* 는 각 손실 항의 가중치이다.

[0078] 이하에서는 각 손실에 대해 상세하게 설명한다.

[0079] $\mathcal{L}_c$ 는 $x_{1:T}^A$ 와 $\hat{x}_{1:T}^A$ 사이의 차이를 최소화하기 위한 사이클 훈련에서의 기본 항이다.

수학식 7

$$[0080] \quad \mathcal{L}_c(x_{1:T}^A, \hat{x}_{1:T}^A) = \|x_{1:T}^A - \hat{x}_{1:T}^A\|_2^2$$

[0081] 비지도 학습 프레임워크에는 리타겟팅된 회전에 대한 명시적인 레이블이 없다. 결과적으로 관절의 과도한 비틀

림이 발생할 수 있어 $\mathcal{L}_t$ 는 관절 회전을 제한한다.

수학식 8

$$\mathcal{L}_t(\hat{q}_{1:T}^B, \hat{q}_{1:T}^A) = \|\max(\mathbf{0}, |\mathcal{E}(\hat{q}_{1:T}^B) - \alpha|)\|_2^2 + \|\max(\mathbf{0}, |\mathcal{E}(\hat{q}_{1:T}^A) - \alpha|)\|_2^2$$

여기서, $\mathcal{E}(\cdot)$ 는 쿼터니언을 회전 각도로 변환한다. α 를 초과하는 각도에 패널티가 부여되며, 본 실시예에서는 $\alpha = 100^\circ$ 이다.

높이(키)가 큰 소스 캐릭터(A)가 보폭을 만들고, 높이가 작은 타겟 캐릭터(B)로 리타게팅하는 것을 가정할 때, B의 모션은 A의 모션보다 작아야 하며 높이 손실은 이러한 효과를 반영한다.

본 실시예에 따르면, 서로 인접한 프레임 사이에서 정의되는 소스 캐릭터의 관절 로컬 좌표의 차이(제1 차이)와 타겟 캐릭터의 관절 로컬 좌표의 차이(제2 차이) 각각을 소스 캐릭터의 높이와 타겟 캐릭터의 높이를 이용하여 정규화하여 계산된 값을 이용하여 높이 손실을 결정한다.

A와 B에서 관절의 로컬 좌표, 즉 $p_{1:T}^A$ 와 $\hat{p}_{1:T}^B$ 를 사용하면 로컬 모션 차이는 인접한 프레임 사이에서 정의된다.

수학식 9

$$d_{2:T}^A = p_{2:T}^A - p_{1:T-1}^A$$

수학식 10

$$\hat{d}_{2:T}^B = \hat{p}_{2:T}^B - \hat{p}_{1:T-1}^B$$

각각 $\mathcal{H}(\bar{s}^A)$ 와 $\mathcal{H}(\bar{s}^B)$ 로 표시되는 A와 B의 높이를 이용하여 크기를 관절 로컬 좌표 차이의 크기를 정규화한다.

수학식 11

$$l_t^A = \frac{\|d_t^A\|}{\mathcal{H}(\bar{s}^A)}$$

수학식 12

$$\hat{l}_t^B = \frac{\|\hat{d}_t^B\|}{\mathcal{H}(\bar{s}^B)}$$

손실은 다음과 같이 정의된다.

수학식 13

[0093] $\text{smooth}_{L_1}(l_{2:T}^A - \hat{l}_{2:T}^B)$

수학식 14

[0094] $\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2, & \text{if } |x| < 1 \\ |x| - 0.5, & \text{otherwise} \end{cases}$

[0095] 여기서, 강건한 L_1 손실은 L_2 손실보다 아웃라이어에 덜 민감하다.

[0096] 보편성의 손실이 없다면, 모션이 글로벌 좌표계의 원점을 중심으로 또는 원점을 가로질러 발생한다고 가정한다. 그러면 소스와 타겟의 루트 관절 회전 정보인 r_t^A 및 $\hat{r}_t^B$ 의 크기를 다음과 같이 정규화할 수 있다.

수학식 15

[0097] $m_t^A = \frac{\|r_t^A\|}{\mathcal{H}(\bar{s}^A)}$

수학식 16

[0098] $\hat{m}_t^B = \frac{\|\hat{r}_t^B\|}{\mathcal{H}(\bar{s}^B)}$

[0099] m_t^A 와 $\hat{m}_t^B$ 를 사용하면, 수학식 13의 손실은 높이 손실을 정의하도록 확장된다.

수학식 17

[0100] $\mathcal{L}_h(x_{1:T}^A, \hat{x}_{1:T}^B) = \text{smooth}_{L_1}(l_{2:T}^A - \hat{l}_{2:T}^B) + \lambda_g \text{smooth}_{L_1}(m_{2:T}^A - \hat{m}_{2:T}^B)$

[0101] 여기서, $\lambda_g = 5$ 이다.

[0102] 인접한 두 프레임 사이의 루트 모션 차이의 정규화된 크기를 고려한다.

수학식 18

[0103] $\delta_{2:T} = \frac{\|r_{2:T} - r_{1:T-1}\|}{\mathcal{H}(\bar{s})}$

[0104] 판별기 D는 A의 실제 모션 시퀀스 및 B의 가상 모션 시퀀스의 스코어를 계산한다.

수학식 19

$$h_{2:T}^A = D(l_{2:T}^A, \delta_{2:T}^A, \mathcal{B}^A)$$

수학식 20

$$h_{2:T}^B = D(\hat{l}_{2:T}^B, \hat{\delta}_{2:T}^B, \mathcal{B}^B)$$

여기서, l_t^A 및 $\hat{l}_t^B$ 는 수학식 11 및 12에 정의되고, $\mathcal{B}^A$ 와 $\mathcal{B}^B$ 는 각각 A와 B의 뼈의 길이를 나타낸다.

훈련 과정에서 뼈대를 무작위로 샘플링 할 때 B는 A와 동일하면, $x_{1:T}^A = \hat{x}_{1:T}^B$ 이다. 적대적 손실은 두 경우를 구별하여 정의한다.

수학식 21

$$\mathcal{L}_a(x_{1:T}^A, \hat{x}_{1:T}^B) = \begin{cases} (h_{2:T}^A)^2 + (1 - h_{2:T}^B)^2 & \text{if } A \neq B \\ \|x_{1:T}^A - \hat{x}_{1:T}^B\|_2^2 & \text{if } A = B \end{cases}$$

여기서, x_t^A 는 p_t^A , q_t^A 및 r_t^A 를 포함하고, $\hat{x}_t^B$ 는 $\hat{p}_t^B$, $\hat{q}_t^B$ 및 $\hat{r}_t^B$ 를 포함한다.

$\mathcal{L}_a$ 는 기본적으로 Villegas의 제안과 두 가지 측면에서 다르다.

하나의 일반적인 GAN에서 시그모이드 교차 엔트로피 손실 함수는 종종 소실 구배 문제를 초래하며, 이를 완화하고 훈련 안정성을 향상시키기 위해 최소 제곱 손실 함수를 사용한다. 다른 하나는 판별기 즉, l_t^A , δ_t^A , $\hat{l}_t^B$ 및 $\hat{\delta}_t^B$ 에 대한 입력은 정규화되며, 위에서 언급한 바와 같이 판별기는 생성기보다 짧은 클립을 사용한다.

본 실시예에서는 단위 쿼터니언에 대한 패널티 항을 사용한다.

수학식 22

$$\mathcal{L}_r(\hat{q}_{1:T}^B, \hat{q}_{1:T}^A) = (1 - \|\hat{q}_{1:T}^B\|)^2 + (1 - \|\hat{q}_{1:T}^A\|)^2$$

상기한 손실은 훈련 안정성을 향상시키는 정규화기 역할을 한다.

높이 손실 $\mathcal{L}_h$ 는 각 관절의 이동 및 변위를 설명하지만 캐릭터의 방향을 다루지는 않는다. θ_t 는 루트 관절의 쿼터니언을 나타낸다. 그리고, 방향 손실 $\mathcal{L}_o$ 는 다음 목적 함수를 최적화한다.

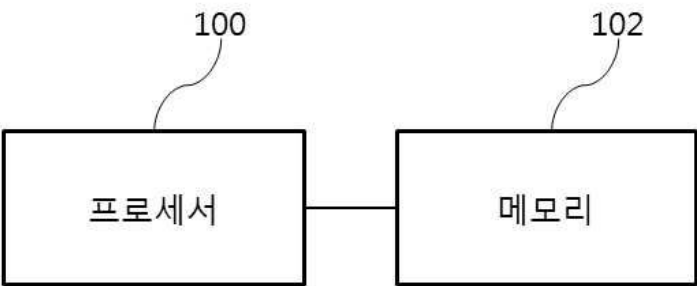
수학식 23

$$\mathcal{L}_o(\theta_{1:T}^A, \hat{\theta}_{1:T}^B) = \text{smooth}_{L_1}(\mathcal{E}(\theta_{1:T}^A) - \mathcal{E}(\hat{\theta}_{1:T}^B))$$

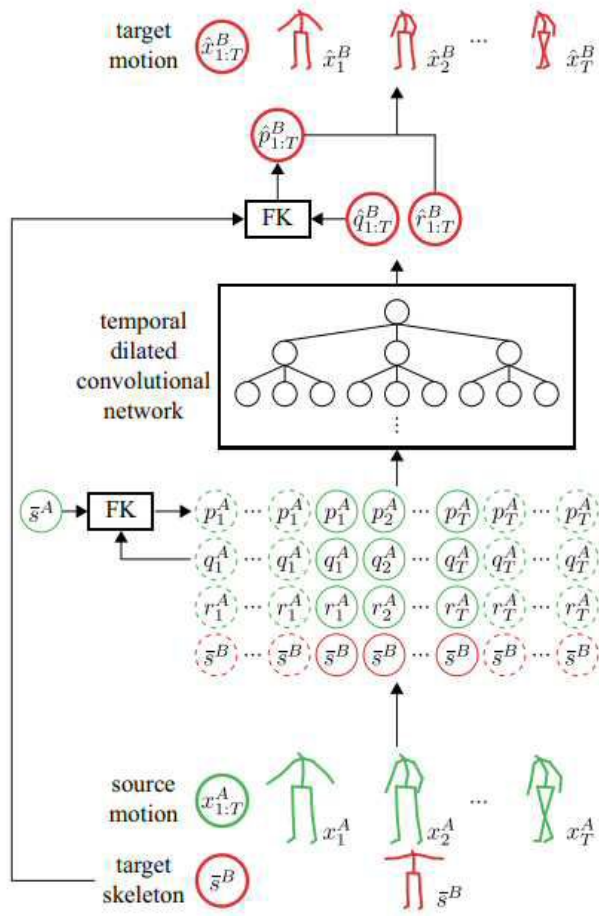
상기한 본 발명의 실시예는 예시의 목적을 위해 개시된 것이고, 본 발명에 대한 통상의 지식을 가지는 당업자라면 본 발명의 사상과 범위 안에서 다양한 수정, 변경, 부가가 가능할 것이며, 이러한 수정, 변경 및 부가는 하기의 특허청구범위에 속하는 것으로 보아야 할 것이다.

도면

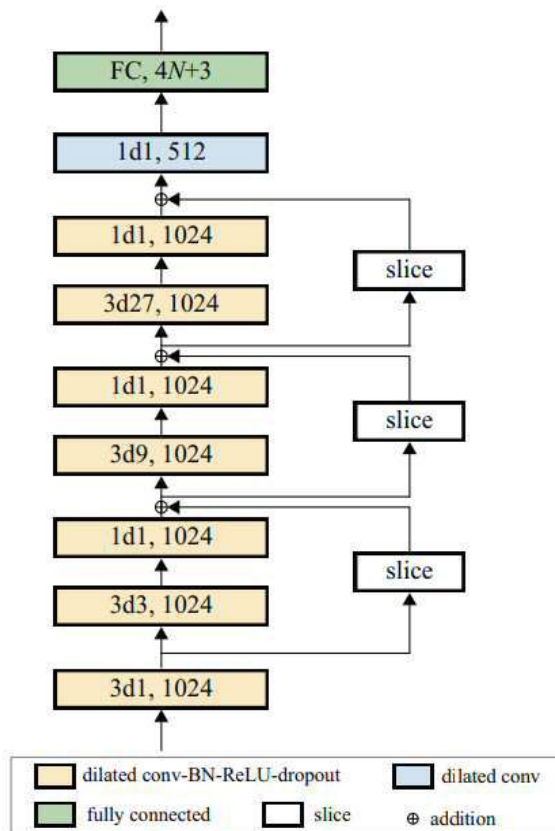
도면1



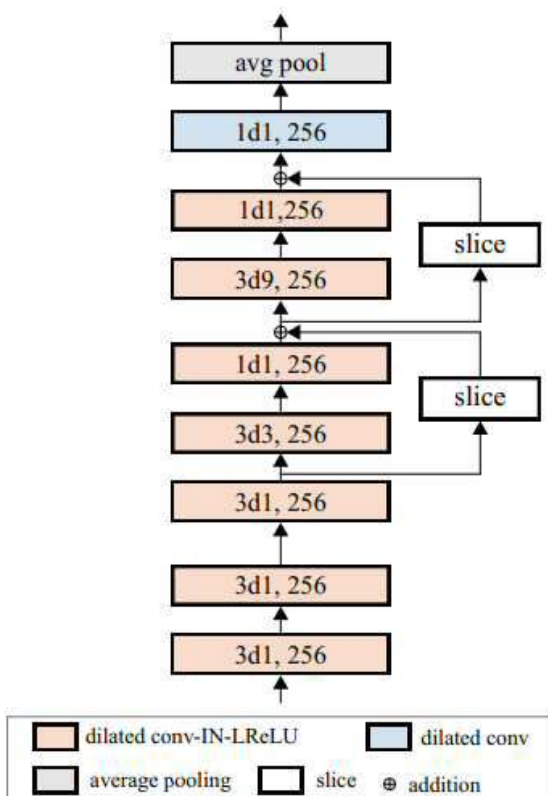
도면2



도면3



도면4



도면5

