(19) 대한민국 지식재산처(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2026년03월26일
(11) 등록번호 10-2941857
(24) 등록일자 2026년03월17일

(51) 국제특허분류(Int. Cl.)
G06T 13/40 (2011.01) *G06F 40/279* (2020.01)
G06N 3/092 (2023.01) *G06T 17/20* (2006.01)
G06T 7/70 (2017.01)
(52) CPC특허분류
G06T 13/40 (2013.01)
G06F 40/279 (2020.01)
(21) 출원번호 10-2025-0076210
(22) 출원일자 2025년06월11일
심사청구일자 2025년06월11일
(56) 선행기술조사문헌
KR102700182 B1*
Lou, Yunhong, et al. "DiverseMotion: Towards diverse human motion generation via discrete diffusion." arXiv preprint arXiv:2309.01372 (2023.09.04.)*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
주식회사 범블비
서울특별시 관악구 행운1길 102, 2층 202호 (봉천동)
(72) 발명자
김대호
서울특별시 관악구 관악로6길 71, 202호
(74) 대리인
특허법인테헤란

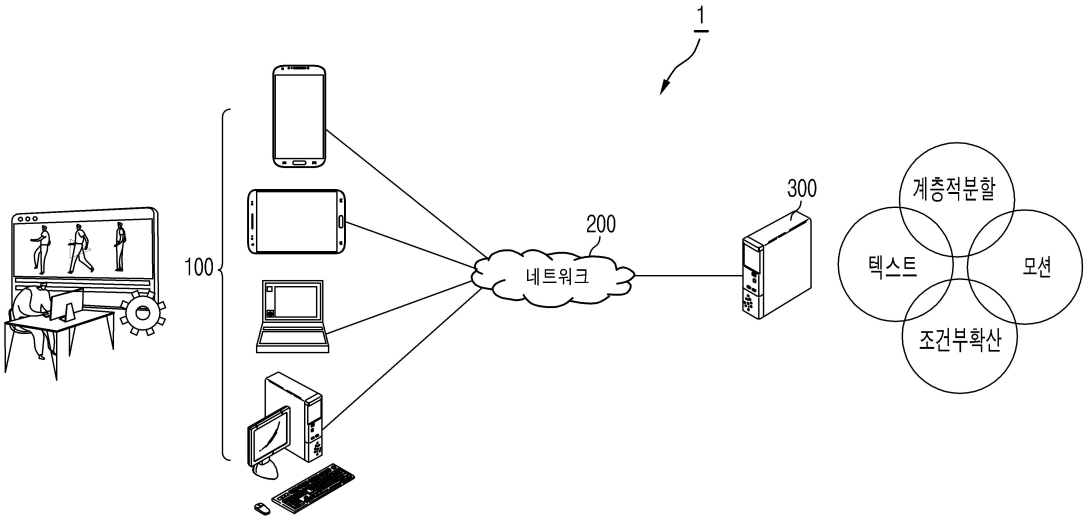
전체 청구항 수 : 총 5 항 심사관 : 전세운

(54) 발명의 명칭 **계층적 확산 기반 텍스트-투-휴머노이드 모션 자동 생성 시스템**

(57) 요 약

계층적 확산 기반 텍스트-투-휴머노이드 모션 자동 생성 시스템이 제공되며, 텍스트를 입력하고, 텍스트에 포함된 모션을 수행하는 3D 객체인 휴머노이드(Humanoid)를 출력하는 사용자 단말 및 사용자 단말로부터 텍스트를 수신하는 수신부, 텍스트에 포함된 동사(Verb)를 추출한 후, 동사를 적어도 하나의 행동 단위로 분해하고, 적어도 하나의 행동 단위의 행동의 순서 및 시간을 추론하는 행동분해부, 적어도 하나의 행동 단위에 대응하는 모션(Motion)을 생성하도록, 기 구축된 조건부 확산 모델(Conditional Diffusion Model)을 호출하여 모션을 생성하는 모션생성부, 적어도 하나의 행동 단위에 대응하는 구간별 모션을 접합하는 접합부를 포함하는 통합제어서버를 포함한다.

대 표 도 - 도1



(52) CPC특허분류
G06N 3/092 (2023.01)
G06T 17/20 (2013.01)
G06T 7/70 (2017.01)
G06T 2207/30196 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	2420015314
과제번호	RS-2024-00514891
부처명	중소벤처기업부
과제관리(전문)기관명	중소기업기술정보진흥원
연구사업명	창업성장기술개발(R&D)
연구과제명	디퓨전 모델 기반 3D 아바타 및 애니메이션 생성 솔루션
과제수행기관명	주식회사 범블비
연구기간	2024.11.01 ~ 2026.10.31

명 세 서

청구범위

청구항 1

텍스트를 입력하고, 상기 텍스트에 포함된 모션을 수행하는 3D 객체인 휴머노이드(Humanoid)를 출력하는 사용자 단말; 및

상기 사용자 단말로부터 텍스트를 수신하는 수신부, 상기 텍스트에 포함된 동사(Verb)를 추출한 후, 상기 동사를 적어도 하나의 행동 단위로 분해하고, 상기 적어도 하나의 행동 단위의 행동의 순서 및 시간을 추론하는 행동분해부, 상기 상기 적어도 하나의 행동 단위에 대응하는 모션(Motion)을 생성하도록, 기 구축된 조건부 확산 모델(Conditional Diffusion Model)을 호출하여 모션을 생성하는 모션생성부, 상기 적어도 하나의 행동 단위에 대응하는 구간별 모션을 접합하는 접합부를 포함하는 통합제어서버;

를 포함하고,

상기 접합부는,

상기 구간별 모션을 접합할 때 겹침 보간을 이용하여 연결하며,

상기 통합제어서버는,

인체를 이루는 적어도 하나의 부위 및 관절의 가동범위, 속도 매끄러움(Velocity Smoothness), 발 접지 제약(Foot Contact Constraint)을 설정 및 구현하는 물리 보정기를, 상기 모션을 생성하는 역확산 과정에 통합시키는 물리보정통합부;

를 더 포함하는 것을 특징으로 하는 계층적 확산 기반 텍스트-투-휴머노이드 모션 자동 생성 시스템.

청구항 2

제 1 항에 있어서,

상기 모션생성부는,

기 설정된 기본 자세에서 시작하고, 상기 기본 자세 이후(Next) 구간은 직전(Previous) 구간의 마지막 프레임을 시작 포즈로 이용하는 것을 특징으로 하는 계층적 확산 기반 텍스트-투-휴머노이드 모션 자동 생성 시스템.

청구항 3

제 1 항에 있어서,

상기 모션생성부는,

상기 행동 단위에 대응하는 모션의 잠재 표현(Latent Features)에, 상기 행동 단위에 대응하는 텍스트의 텍스트 임베딩(Embedding)을, 상기 조건부 확산 모델(Conditional Diffusion Model)에 교차 어텐션(Cross-Attention) 또는 결합(Concatenation)으로 주입하는 것을 특징으로 하는 계층적 확산 기반 텍스트-투-휴머노이드 모션 자동 생성 시스템.

청구항 4

삭제

청구항 5

삭제

청구항 6

제 1 항에 있어서,

상기 물리 보정기는,

규칙 기반(Rule-Based) 또는 강화학습 기반(Reinforcement Learning Based)의 물리 보정기이고,

상기 물리 보정기는 상기 역확산 과정에 통합할 수 있도록 엔드-투-엔드(End-to-End) 구조로 동작되는 것을 특징으로 하는 계층적 확산 기반 텍스트-투-휴머노이드 모션 자동 생성 시스템.

청구항 7

제 1 항에 있어서,

상기 통합제어서버는,

생성된 모션을 기 설정된 3D 모델링 포맷으로 출력하고, 모션 캡셔닝을 통하여 상기 텍스트 및 모션 간 일치를 평가하며, 상기 일치 여부에 따라 자동 피드백 루프를 구동시키는 출력검증부;

를 더 포함하는 것을 특징으로 하는 계층적 확산 기반 텍스트-투-휴머노이드 모션 자동 생성 시스템.

발명의 설명

기 술 분 야

[0001] 본 발명은 계층적 확산 기반 텍스트-투-휴머노이드 모션 자동 생성 시스템에 관한 것으로, 텍스트에 포함된 동작을 행동 단위로 분해하고, 행동 단위에 대응하는 모션을 조건부 확산 모델로 생성한 후, 각 모션을 시계열적으로 이어붙이는 시스템을 제공한다.

배 경 기 술

[0002] 최근 인공지능 기술의 발전과 함께, 자연어 설명으로부터 사람의 동작을 자동 생성하는 텍스트-투-모션(Text-to-Motion) 기술이 게임, 애니메이션, 가상 인간, 로봇 제어 등 다양한 분야에서 높은 관심을 받고 있다. 특히 2022년 이후 확산 모델(Diffusion Model)의 급격한 성능 향상은 이미지 생성 뿐 아니라, 3D 휴머노이드 모션 생성 영역에서도 확산 기반 접근의 활용 가능성을 크게 확대시켰다. 확산 모델은 기존의 GAN(Generative Adversarial Network)에 비해 학습이 안정적이고 생성 다양성이 높다는 강점으로 인해, MotionDiffuse, MDM(Motion Diffusion Model) 등 다양한 기법에 적용되어 왔다.

[0003] 이때, 텍스트 기반으로 모션을 생성하는 방법이 연구 및 개발되었는데, 이와 관련하여, 선행기술인 한국공개특허 제2025-0053700호(2025년04월22일 공개) 및 한국공개특허 제2024-0147351호(2024년10월08일 공개)에는, 자연어 문장에서 동작을 추출하고, 이를 조건으로 모션 시퀀스를 생성할 때, RNN 및 VAE 기반의 시계열 예측 모델을 이용하여 사람의 움직임을 생성하며, 생성된 모션은 일정한 형식으로 출력되어 3D 캐릭터에 적용되는 구성과, 입력 문장의 텍스트에 텍스트 임베딩을 이용하여 GAN 기반 모델로 모션을 생성하는 구성이 각각 개시되어 있다.

[0004] 다만. 전자의 경우, 물리적 제약이나 세밀한 동작 표현보다는 전반적인 동작 흐름 생성에 초점이 맞춰져 있고, 후자의 경우, 생성된 모션의 연속성 및 물리적 일관성 확보에는 상대적으로 취약하며, 시간적으로 긴 시퀀스를 다루는 데 한계가 존재한다. 또, 현재의 텍스트-투-모션 기술은 여전히 여러 근본적인 제약에 직면하고 있는데, 학습 데이터가 절대적으로 부족하고, 시간축 상의 물리적 연속성과 생체 구조적 타당성을 함께 만족해야 한다는 점에서, 사람의 실제 움직임을 정밀하게 생성하기 위한 난이도가 매우 높다는 것이다. 이에, 확산 모델 기반 텍스트-투-모션 기술이 직면한 한계를 극복할 수 있는 시스템의 연구 및 개발이 요구된다.

발명의 내용

해결하려는 과제

[0005] 본 발명의 일 실시예는, 사용자 단말로부터 텍스트가 입력되면 행동 단위로 분할하고, 행동 단위에 대응하는 모

션을 생성하도록 조건부 확산 모델을 이용하며, 생성된 각 모션이 자연스럽게 연결될 수 있도록 후처리를 함으로써, 생성가능한 동작길이의 제한을 없애고, 다중액션 및 순차 동작처리의 한계를 극복하며, 텍스트의 조건을 명확히 반영한 모션을 생성할 수 있도록 하는, 계층적 확산 기반 텍스트-투-휴머노이드 모션 자동 생성 시스템을 제공할 수 있다. 다만, 본 실시예가 이루고자 하는 기술적 과제는 상기된 바와 같은 기술적 과제로 한정되지 않으며, 또 다른 기술적 과제들이 존재할 수 있다.

과제의 해결 수단

[0006] 상술한 기술적 과제를 달성하기 위한 기술적 수단으로서, 본 발명의 일 실시예는, 텍스트를 입력하고, 텍스트에 포함된 모션을 수행하는 3D 객체인 휴머노이드(Humanoid)를 출력하는 사용자 단말 및 사용자 단말로부터 텍스트를 수신하는 수신부, 텍스트에 포함된 동사(Verb)를 추출한 후, 동사를 적어도 하나의 행동 단위로 분해하고, 적어도 하나의 행동 단위의 행동의 순서 및 시간을 추론하는 행동분해부, 적어도 하나의 행동 단위에 대응하는 모션(Motion)을 생성하도록, 기 구축된 조건부 확산 모델(Conditional Diffusion Model)을 호출하여 모션을 생성하는 모션생성부, 적어도 하나의 행동 단위에 대응하는 구간별 모션을 접합하는 접합부를 포함하는 통합제어서버를 포함한다.

발명의 효과

[0007] 전술한 본 발명의 과제 해결 수단 중 어느 하나에 의하면, 사용자 단말로부터 텍스트가 입력되면 행동 단위로 분할하고, 행동 단위에 대응하는 모션을 생성하도록 조건부 확산 모델을 이용하며, 생성된 각 모션이 자연스럽게 연결될 수 있도록 후처리를 함으로써, 생성가능한 동작길이의 제한을 없애고, 다중액션 및 순차 동작처리의 한계를 극복하며, 텍스트의 조건을 명확히 반영한 모션을 생성할 수 있도록 한다.

도면의 간단한 설명

[0008] 도 1은 본 발명의 일 실시예에 따른 계층적 확산 기반 텍스트-투-휴머노이드 모션 자동 생성 시스템을 설명하기 위한 도면이다.

도 2는 도 1의 시스템에 포함된 통합제어서버를 설명하기 위한 블록 구성도이다.

도 3 및 도 4는 본 발명의 일 실시예에 따른 모션 자동 생성 서비스가 구현된 일 실시예를 설명하기 위한 도면이다.

도 5는 본 발명의 일 실시예에 따른 모션 자동 생성 서비스 제공 방법을 설명하기 위한 동작 흐름도이다.

발명을 실시하기 위한 구체적인 내용

[0009] 아래에서는 첨부한 도면을 참조하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 본 발명의 실시예를 상세히 설명한다. 그러나 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시예에 한정되지 않는다. 그리고 도면에서 본 발명을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.

[0010] 명세서 전체에서, 어떤 부분이 다른 부분과 "연결"되어 있다고 할 때, 이는 "직접적으로 연결"되어 있는 경우뿐 아니라, 그 중간에 다른 소자를 사이에 두고 "전기적으로 연결"되어 있는 경우도 포함한다. 또한 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미하며, 하나 또는 그 이상의 다른 특징이나 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

[0011] 명세서 전체에서 사용되는 정도의 용어 "약", "실질적으로" 등은 언급된 의미에 고유한 제조 및 물질 허용오차가 제시될 때 그 수치에서 또는 그 수치에 근접한 의미로 사용되고, 본 발명의 이해를 돕기 위해 정확하거나 절대적인 수치가 언급된 개시 내용을 비양심적인 침해자가 부당하게 이용하는 것을 방지하기 위해 사용된다. 본 발명의 명세서 전체에서 사용되는 정도의 용어 "~(하는) 단계" 또는 "~의 단계"는 "~ 를 위한 단계"를 의미하지 않는다.

[0012] 본 명세서에 있어서 '부(部)'란, 하드웨어에 의해 실현되는 유닛(unit), 소프트웨어에 의해 실현되는 유닛, 양방을 이용하여 실현되는 유닛을 포함한다. 또한, 1 개의 유닛이 2 개 이상의 하드웨어를 이용하여 실현되어도

되고, 2 개 이상의 유닛이 1 개의 하드웨어에 의해 실현되어도 된다. 한편, '~부'는 소프트웨어 또는 하드웨어에 한정되는 의미는 아니며, '~부'는 어드레싱 할 수 있는 저장 매체에 있도록 구성될 수도 있고 하나 또는 그 이상의 프로세서들을 재생시키도록 구성될 수도 있다. 따라서, 일 예로서 '~부'는 소프트웨어 구성요소들, 객체 지향 소프트웨어 구성요소들, 클래스 구성요소들 및 태스크 구성요소들과 같은 구성요소들과, 프로세스들, 함수들, 속성들, 프로시저들, 서브루틴들, 프로그램 코드의 세그먼트들, 드라이버들, 펌웨어, 마이크로코드, 회로, 데이터, 데이터베이스, 데이터 구조들, 테이블들, 어레이들 및 변수들을 포함한다. 구성요소들과 '~부'들 안에서 제공되는 기능은 더 작은 수의 구성요소들 및 '~부'들로 결합되거나 추가적인 구성요소들과 '~부'들로 더 분리될 수 있다. 뿐만 아니라, 구성요소들 및 '~부'들은 디바이스 또는 보안 멀티미디어카드 내의 하나 또는 그 이상의 CPU들을 재생시키도록 구현될 수도 있다.

[0013] 본 명세서에 있어서 단말, 장치 또는 디바이스가 수행하는 것으로 기술된 동작이나 기능 중 일부는 해당 단말, 장치 또는 디바이스와 연결된 서버에서 대신 수행될 수도 있다. 이와 마찬가지로, 서버가 수행하는 것으로 기술된 동작이나 기능 중 일부도 해당 서버와 연결된 단말, 장치 또는 디바이스에서 수행될 수도 있다.

[0014] 본 명세서에서 있어서, 단말과 매핑(Mapping) 또는 매칭(Matching)으로 기술된 동작이나 기능 중 일부는, 단말의 식별 정보(Identifying Data)인 단말기의 고유번호나 개인의 식별정보를 매핑 또는 매칭한다는 의미로 해석될 수 있다.

[0015] 이하 첨부된 도면을 참고하여 본 발명을 상세히 설명하기로 한다.

[0016] 도 1은 본 발명의 일 실시예에 따른 계층적 확산 기반 텍스트-투-휴머노이드 모션 자동 생성 시스템을 설명하기 위한 도면이다. 도 1을 참조하면, 계층적 확산 기반 텍스트-투-휴머노이드 모션 자동 생성 시스템(1)은, 적어도 하나의 사용자 단말(100) 및 통합제어서버(300)을 포함할 수 있다. 다만, 이러한 도 1의 계층적 확산 기반 텍스트-투-휴머노이드 모션 자동 생성 시스템(1)은, 본 발명의 일 실시예에 불과하므로, 도 1을 통하여 본 발명이 한정 해석되는 것은 아니다.

[0017] 이때, 도 1의 각 구성요소들은 일반적으로 네트워크(Network, 200)를 통해 연결된다. 예를 들어, 도 1에 도시된 바와 같이, 적어도 하나의 사용자 단말(100)은 네트워크(200)를 통하여 통합제어서버(300)와 연결될 수 있다. 그리고, 통합제어서버(300)는, 네트워크(200)를 통하여 적어도 하나의 사용자 단말(100)과 연결될 수 있다.

[0018] 여기서, 네트워크는, 복수의 단말 및 서버들과 같은 각각의 노드 상호 간에 정보 교환이 가능한 연결 구조를 의미하는 것으로, 이러한 네트워크의 일 예에는 근거리 통신망(LAN: Local Area Network), 광역 통신망(WAN: Wide Area Network), 인터넷(WWW: World Wide Web), 유무선 데이터 통신망, 전화망, 유무선 텔레비전 통신망 등을 포함한다. 무선 데이터 통신망의 일례에는 3G, 4G, 5G, 3GPP(3rd Generation Partnership Project), 5GPP(5th Generation Partnership Project), 5G NR(New Radio), 6G(6th Generation of Cellular Networks), LTE(Long Term Evolution), WIMAX(World Interoperability for Microwave Access), 와이파이(Wi-Fi), 인터넷(Internet), LAN(Local Area Network), Wireless LAN(Wireless Local Area Network), WAN(Wide Area Network), PAN(Personal Area Network), RF(Radio Frequency), 블루투스(Bluetooth) 네트워크, NFC(Near-Field Communication) 네트워크, 위성 방송 네트워크, 아날로그 방송 네트워크, DMB(Digital Multimedia Broadcasting) 네트워크 등이 포함되나 이에 한정되지는 않는다.

[0019] 하기에서, 적어도 하나의 라는 용어는 단수 및 복수를 포함하는 용어로 정의되고, 적어도 하나의 라는 용어가 존재하지 않더라도 각 구성요소가 단수 또는 복수로 존재할 수 있고, 단수 또는 복수를 의미할 수 있음은 자명하다 할 것이다. 또한, 각 구성요소가 단수 또는 복수로 구비되는 것은, 실시예에 따라 변경가능하다 할 것이다.

[0020] 적어도 하나의 사용자 단말(100)은, 모션 자동 생성 서비스 관련 웹 페이지, 앱 페이지, 프로그램 또는 애플리케이션을 이용하여 텍스트를 입력하고, 텍스트에 대응하는 모션이 휴머노이드에서 구현되는 결과물을 출력하는 사용자(User)의 단말일 수 있다.

[0021] 여기서, 적어도 하나의 사용자 단말(100)은, 네트워크를 통하여 원격지의 서버나 단말에 접속할 수 있는 컴퓨터로 구현될 수 있다. 여기서, 컴퓨터는 예를 들어, 내비게이션, 웹 브라우저(WEB Browser)가 탑재된 노트북, 데스크톱(Desktop), 랩톱(Laptop) 등을 포함할 수 있다. 이때, 적어도 하나의 사용자 단말(100)은, 네트워크를 통해 원격지의 서버나 단말에 접속할 수 있는 단말로 구현될 수 있다. 적어도 하나의 사용자 단말(100)은, 예를 들어, 휴대성과 이동성이 보장되는 무선 통신 장치로서, 내비게이션, PCS(Personal Communication System),

GSM(Global System for Mobile communications), PDC(Personal Digital Cellular), PHS(Personal Handyphone System), PDA(Personal Digital Assistant), IMT(International Mobile Telecommunication)-2000, CDMA(Code Division Multiple Access)-2000, W-CDMA(W-Code Division Multiple Access), Wibro(Wireless Broadband Internet) 단말, 스마트폰(Smartphone), 스마트 패드(Smartpad), 타블렛 PC(Tablet PC) 등과 같은 모든 종류의 핸드헬드(Handheld) 기반의 무선 통신 장치를 포함할 수 있다.

[0022] 통합제어서버(300)는, 모션 자동 생성 서비스 웹 페이지, 앱 페이지, 프로그램 또는 애플리케이션을 제공하는 서버일 수 있다. 그리고, 통합제어서버(300)는, 사용자 단말(100)에서 텍스트를 수신하고, 동사를 추출하여 행동 단위로 분할하며, 행동 단위에 대응하는 모션을 생성하도록 기 구축된 조건부 확산 모델을 호출하고, 조건부 확산 모델에서 생성된 모션을 구간별로 접합해서 휴머노이드에서 일련의(A Series of) 모션을 수행하도록 하는 서버일 수 있다. 여기서, 통합제어서버(300)는, 네트워크를 통하여 원격지의 서버나 단말에 접속할 수 있는 컴퓨터로 구현될 수 있다. 여기서, 컴퓨터는 예를 들어, 내비게이션, 웹 브라우저(WEB Browser)가 탑재된 노트북, 데스크톱(Desktop), 랩톱(Laptop) 등을 포함할 수 있다.

[0023] 도 2는 도 1의 시스템에 포함된 통합제어서버를 설명하기 위한 블록 구성도이고, 도 3 및 도 4는 본 발명의 일 실시예에 따른 모션 자동 생성 서비스가 구현된 일 실시예를 설명하기 위한 도면이다.

[0024] 도 2를 참조하면, 통합제어서버(300)는, 수신부(310), 행동분해부(320), 모션생성부(330), 접합부(340), 물리보정통합부(350) 및 출력검증부(360)를 포함할 수 있다.

[0025] 본 발명의 일 실시예에 따른 통합제어서버(300)나 연동되어 동작하는 다른 서버(미도시)가 적어도 하나의 사용자 단말(100)로 모션 자동 생성 서비스 애플리케이션, 프로그램, 앱 페이지, 웹 페이지 등을 전송하는 경우, 적어도 하나의 사용자 단말(100)은, 모션 자동 생성 서비스 애플리케이션, 프로그램, 앱 페이지, 웹 페이지 등을 설치하거나 열 수 있다. 또한, 웹 브라우저에서 실행되는 스크립트를 이용하여 서비스 프로그램이 적어도 하나의 사용자 단말(100)에서 구동될 수도 있다. 여기서, 웹 브라우저는 웹(WWW: World Wide Web) 서비스를 이용할 수 있게 하는 프로그램으로 HTML(Hyper Text Mark-up Language)로 서술된 하이퍼 텍스트를 받아서 보여주는 프로그램을 의미하며, 예를 들어 크롬(Chrome), 에지(Microsoft Edge), 사파리(Safari), 파이어폭스(FireFox), 웨일(Whale), UC 브라우저 등을 포함한다. 또한, 애플리케이션은 단말 상의 응용 프로그램(Application)을 의미하며, 예를 들어, 모바일 단말(스마트폰)에서 실행되는 앱(App)을 포함한다.

[0026] 도 2를 참조하면, 수신부(310)는, 사용자 단말(100)로부터 텍스트를 수신할 수 있다. 사용자 단말(100)은 텍스트를 입력할 수 있다. 여기서, 텍스트는 자연어로 이루어질 수도 있고, 문장 형태일 수도 있으며, 키워드 중심의 형태일 수도 있다.

[0027] <행동분해>

[0028] 행동분해부(320)는, 텍스트에 포함된 동사(Verb)를 추출한 후, 동사를 적어도 하나의 행동 단위로 분해하고, 적어도 하나의 행동 단위(Action Unit)의 행동의 순서 및 시간을 추론할 수 있다. 이를 위하여, 행동분해부(320)는, ① 우선 입력된 텍스트를, 형태소 분석 및 구문 분석을 통해 행동 단위로 분해해야 한다. 또, ② 각 행동 단위의 행동의 앞뒤 관계를 분석해서 순서를 설정하고, ③ 각 행동 단위의 행동이 얼마나 지속될 것인지를 추정해야 한다. 그리고, ④ 문장이 복합문장인 경우 상위문장-하위문장으로 구분하여 계층적 시나리오 트리구조를 추출해야 한다.

표 1

[0029]

① 문장 분해	
입력된 복합 텍스트 문장을 동사 기반의 행동 단위(Action Unit)로 분해	
1. 형태소 분석 및 품사 태깅	-한국어: KoNLPy(Mecab), 영어: SpaCy, Stanza 등 사용 -동사(Verb), 부사(Adverb), 연결어(Conjunction) 등을 추출하여 문장 구조 파악
2. 구문 분석 (Syntactic Parsing)	-문장의 Dependency Tree 또는 Constituency Tree를 생성 -예: “사람이 점프한 뒤 착지하여 달린다” → 동작 순서를 구성하는 핵심 동사 노드를 추출: ▶점프하다 → 착지하다 → 달리다

3. 행동 단위 결정	-각 동사에 대해 해당 행동이 모션 생성 가능한 최소 단위인지 확인 -불명확한 동사 또는 복합동사는 기 정의된 행동사전(Action Ontology)과 매핑하여 표준동작으로 변환 -예: “넘어진다” → “균형 잃음 + 낙하 + 착지”

표 2

[0030]

② 행동 순서 결정	
문장에서 추출한 행동 단위들의 시계열적 순서를 정하는 것	
1. 시간적 접속어/표현 분석	-예: “그리고”, “이후에”, “먼저”, “~한 뒤에” 등을 분석해 순차성 추론 -문장 내 연결어의 위치 기준으로 행동 순서를 지정
2. 의미적 시간 관계 추론	-GPT, T5, BERT 기반의 시퀀스 순서 예측 모델을 통해 내재된 시간 순서를 추론 -문장이 직접적 시점 표현이 없을 경우 문맥적 상식 기반 순서 결정
3. 타임라인(Time Line) 상 정렬	-정해진 행동 리스트를 순차적으로 타임라인 상에 배치 -예: [점프(t=0~1.5s)] → [착지(1.5~2.5s)] → [달리기(2.5~5s)]

표 3

[0031]

③ 예상 지속 시간 추정	
각 행동이 실제로 몇 초간 지속될지 예측	
1. 동작-시간 쌍(Pair) 데이터셋 기반 회귀 모델 학습	-예: HumanML3D 같은 모션 데이터셋에서 -(텍스트: “점프한다”, 실제 모션 길이: 1.2초) -(텍스트: “걷는다”, 실제 모션 길이: 4.0초) 와 같은 쌍으로 회귀모델 학습
2. 예상 지속시간 추론 모델	-행동 단위별 텍스트 임베딩을 입력으로 하는 회귀 예측 모델(e.g., BERT + MLP head) 사용 -모델은 문맥에 따라 “점프”는 1초, “달리기”는 3초 등 동작 특성 기반으로 예측
3. 복합 문장인 경우 비율 분할	-“점프하고 착지한 후 달린다(5초)”처럼 전체 시퀀스 시간이 고정되면, 각 구간별 상대 비율 기반으로 배분 가능 → 점프: 20%, 착지: 20%, 달리기: 60%

[0032] 즉 전체적인 흐름을 요약하면, 첫 번째로 형태소 및 구문 분석으로 핵심 동작의 동사를 추출하고, 두 번째로 문법적 및 의미적 시간관계 기반으로 정렬하여 시계열 순서를 추론하며, 세 번째로 학습 기반 회귀모델 또는 룰 기반으로 지속 시간을 예측할 수 있다. 이러한 과정을 거치면 입력된 텍스트는 이하 표 4와 같은 구조로 변환되게 된다.

표 4

[0033]

```
[
  {"동작": "점프", "순서": 1, "시간": 1.5},
  {"동작": "착지", "순서": 2, "시간": 1.0},
  {"동작": "달리기", "순서": 3, "시간": 3.5}
]
```

[0034] 이 구조는 이후 모션생성모델인 조건부 확산 모델의 입력으로 사용되어, 각 행동 단위에 맞는 모션을 생성하는 기반이 된다.

[0035] <모션생성>

[0036] 모션생성부(330)는, 적어도 하나의 행동 단위에 대응하는 모션(Motion)을 생성하도록, 기 구축된 조건부 확산 모델(Conditional Diffusion Model)을 호출하여 모션을 생성할 수 있다. 모션생성부(330)는, 기 설정된 기본

자세에서 시작하고, 기본 자세 이후(Next) 구간은 직전(Previous) 구간의 마지막 프레임을 시작 포즈로 이용할 수 있다. 이때, 기본 자세란, T-포즈 즉 사람이 양팔을 벌리고 선 포즈를 의미한다. 예를 들어, 행동 단위가 [점프 → 착지 → 달리기]였다면, 각 행동 단위, 즉 점프, 착지 및 달리기 각각에 대하여 조건부 확산 모델을 적용하여 점프, 착지 및 달리기 각각의 모션을 각각 생성할 수 있다.

[0037] 이때, 확산 모델은 랜덤한 노이즈에서 시작해서 점점 의미있는 데이터를 복원해내는 방식이다. 이미지 생성에서 유명한 스테이블 디퓨전(Stable Diffusion)처럼, 랜덤한 점들에서 출발해 점차 실제같은 결과물을 만들어낸다. 확산 모델의 핵심 아이디어는, 데이터에 점점 노이즈를 추가해 없애버린 후, 그 노이즈를 거꾸로 제거하면서 원래 데이터를 복원하는 방식인데, 예를 들어, 사람의 걸어가는 모션이 있다고 가정하면, 이 모션에 점점 노이즈를 섞어서 이상한 움직임을 만들고, 다시 거꾸로 그 노이즈를 제거하는 방법을 학습해서, 나중에는 완전히 노이즈(랜덤값)에서 시작해도 사람의 걸어가는 모션을 만들 수 있도록 하는 것이다. GAN및 확산 모델 둘 다 모두 생성 모델(Generative Model)에 속하지만 작동 방식은 이하 표 5와 같이 다르다.

표 5

[0038]

항목	GAN	확산 모델
작동 방식	생성자 vs 판별자 간의 게임	노이즈 → 점진적 제거(역확산)
훈련 방식	경쟁 훈련(Adversarial)	자기 자신을 예측(Denoising Objective)
대표 모델	StyleGAN, BigGAN 등	DDPM, Stable Diffusion, Imagen 등
장점	빠른 생성 속도	안정적 훈련, 높은 품질
단점	모드 붕괴, 훈련 불안정	생성 속도 느림(많은 스텝 필요)

[0039] 이러한 확산 모델이 본 발명의 모션 생성에 쓰인 이유는, 모션은 시계열 구조(프레임 연속성)가 중요한데, GAN은 훈련이 불안정하고 연속적인 물리적 자연스러움이 깨지기 쉽다. 반면, 확산 모델은 연속성 보존, 세부 묘사 정확성에 강해서 유리하다. 이에, 텍스트-투-모션 분야에서는, 이미지 생성에 쓰이던 확산 모델을 변형해서 사람의 움직임 시퀀스를 생성하는 데 사용하는 것이 바람직하다. 이때, 조건부 확산 모델이란, 단순히 무작위 데이터를 생성하는 일반적인 확산 모델과 달리, 특정 조건(Condition)에 맞는 데이터를 생성하는 확산 모델인데, 일반적인 확산 모델과 비교하면 이하 표 6과 같다.

표 6

[0040]

일반 확산 모델(Unconditional Diffusion Model)	조건부 확산 모델(Conditional Diffusion Model)
-랜덤 노이즈에서 시작해서 데이터(예: 이미지, 모션 등)를 점진적으로 복원 -아무 조건 없이 그냥 “그럴듯한” 샘플 생성	-텍스트, 이미지, 포즈 등 외부 조건을 함께 입력하여, -그 조건에 적합한 결과물을 생성

[0041] 조건부 확산 모델은, 예를 들어, "점프하는 사람"이라는 텍스트를 조건으로 넣으면, 점프하는 모션을 생성하게 된다.작동 구조는 첫 번째로 정방향 노이즈 추가(Forward Process)를 하면서 원래 모션을 점진적으로 노이즈로 오염시키고, 두 번째로, 역방향 복원(Reverse Process)을 통하여, 완전한 노이즈 상태에서 시작해서 점차 원래 모션으로 복원하게 된다. 즉, 텍스트 임베딩 등 조건 정보(Condition)을 사용하여 “그 조건에 맞는 모션”으로 유도하는 것이다.

[0042] 모션생성부(330)는, 행동 단위에 대응하는 모션의 잠재 표현(Latent Features)에, 행동 단위에 대응하는 텍스트의 텍스트 임베딩(Embedding)을, 조건부 확산 모델(Conditional Diffusion Model)에 교차 어텐션(Cross-Attention) 또는 결합(Concatenation)으로 주입할 수 있다. 우선 텍스트가 입력되면, CLIP 등의 사전학습(Pre-Trained)모델로 텍스트 임베딩 생성하고, 포즈 시퀀스 입력 또는 노이즈 입력이 입력되면, Transformer-U-Net이 모션 잠재 표현(Latent Feature)을 생성한다. 이 잠재 표현은 텍스트 임베딩과 결합되어 조건부 확률 모델로 입력된다. 데이터의 결합 방식은 이하 표 7과 같은 방식 중 하나일 수 있다.

표 7

[0043]

방식	설명
결합(Concatenation)	노이즈 입력에 조건 벡터를 단순히 이어붙임
크로스-어텐션(Cross-Attention)	Transformer 구조에서 조건을 ‘질문’처럼 사용하여 노이즈를 조정

Classifier-Free Guidance	조건을 적용한 버전과 적용하지 않은 버전을 동시에 학습하여 유연하게 조정

[0044] 이때 조건부 확산 모델은, Transformer-U-Net 기반 아키텍처를 이용할 수 있는데, 이는 시간축 데이터(시퀀스)에 잘 맞도록, U-Net 구조의 확산 모델에 Transformer 블록을 결합한 형태의 아키텍처를 의미한다. 이때, Transformer-U-Net 기반 아키텍처는, 확산 과정에서 노이즈를 데이터로 복원하는 네트워크 구조를 의미한다. 예를 들어, 텍스트가 [걷다가 점프함]이라면, 텍스트 임베딩(CLIP 등)을 포즈 정보(모션 시퀀스의 잠재 표현(Latent Features))와 결합(Concatenation)한 결과를, Transformer-U-Net에 조건으로 주입하면, Transformer-U-Net 기반 아키텍처를 가진 조건부 확산 모델은, 완전한 노이즈에서 모션 생성을 시작하다가, 점차 노이즈를 제거하고, "걷다→ 점프하다"에 해당하는 모션 시퀀스를 생성할 수 있다.

[0045] <모션접합>

[0046] 접합부(340)는, 적어도 하나의 행동 단위에 대응하는 구간별 모션을 접합할 수 있다. 이때, [행동 단위 → 모션 → 구간]의 관계는, [의미 → 표현 → 시간]이라고 표현할 수 있다. 행동 단위인 동작을 표현한 것이 모션이고, 이 모션이 일어나는 시퀀스를 하나의 구간이라고 명명할 수 있다. 예를 들어, 텍스트 입력이 [사람이 점프한 후 착지하고, 달리기 시작한다]였고, 행동 단위가 [점프] → [착지] → [달리기]로 분리(분해, 분할)되었으며, 각 행동 단위에 대하여, 각 모션 및 지속 시간이, [점프 모션-약 1.2초], [착지 모션-약 0.8초], [달리기 모션-약 3.0초]으로 추론된 경우, 구간이란, [점프: t=0.0~1.2], [착지: t=1.2~2.0], [달리기: t=2.0~5.0]와 같이 시간축 상의 표현이다. 즉, 생성된 모션이 전체 시퀀스 내에서 차지하는 시간적 범위를 의미한다.

표 8

[0047]

용어	정의
행동 단위(Action Unit)	텍스트 시나리오에서 추출한 의미 있는 최소 동작 단위(예: 점프, 착지, 걷기)
모션(Motion)	각 행동 단위에 해당하는 구체적인 시계열 움직임 데이터(예: 3초간의 점프 동작)
구간(Segment)	하나의 모션이 차지하는 시간적 범위(예: t=3.0초 ~ 5.0초); 일반적으로 행동 단위 1개 = 구간 1개

[0048] 이렇게 구간별 모션이 각각 생성되었다면, 이 각각의 모션을 이제 이어붙여야 한다. 이에, 접합부(340)는, 구간별 모션을 접합할 때 겹침 보간을 이용하여 연결할 수 있다. 이를 위하여 구간별 모션을 SLERP(Spherical Linear Interpolation) 등의 선형 보간(Linear Interpolation) 기법으로 부드럽게 연결하게 된다. SLERP는, 회전(Interpolation) 사이를 부드럽게 연결하는 보간 기법으로, 3D 회전(쿼터니언)에 주로 사용되는 기법이다. 이를 통하여 구간별 생성된 모션을 부드럽게 연결해 이음매를 최소화할 수 있다.

[0049] 물리보정통합부(350)는, 인체를 이루는 적어도 하나의 부위 및 관절의 가동범위, 속도 매끄러움(Velocity Smoothness), 발 접지 제약(Foot Contact Constraint)을 설정 및 구현하는 물리 보정기를, 모션을 생성하는 역확산 과정에 통합시킬 수 있다. 이때, 속도 매끄러움(Velocity Smoothness)은 모션 생성 시스템에서 시간에 따른 관절 또는 루트의 움직임(속도)이 갑작스럽게 튀거나 멈추지 않도록 제어하는 개념이다. 즉, 모션의 부드러움을 수치적으로 보장하기 위한 물리적 제약 조건이다.

[0050] 또, 발 접지 제약은, 모션 시퀀스 중 발이 땅에서 불필요하게 떠 있거나, 미끄러지거나, 관통하지 않도록 강제하는 규칙 또는 손실 함수를 의미한다. 실제로 인간이 움직일 때, 발이 바닥에서 갑자기 붕 뜨거나, 착지한 발이 미끄러지거나, 제자리 걸음인데 발이 계속 움직이는 현상은 물리적으로 말이 되지 않는다. 하지만 AI로 모션을 생성할 경우, 이런 부자연스러운 동작이 자주 발생하게 되므로, 이러한 물리 법칙을 적용하고 부자연스러운 동작을 막기 위하여 "발 접지 제약"이 필요하다.

[0051] 이러한 물리 보정기는, 규칙 기반(Rule-Based) 또는 강화학습 기반(Reinforcement Learning Based)의 물리 보정기일 수 있다. 예를 들어, 규칙 기반 모델 또는 강화학습 기반 모델을 통하여, "이전 프레임의 속도와 차이가 너무 큰 경우 → 페널티 부여" 또는 속도가 급변할 경우 보간 방식으로 자동 수정 등의 방식을 적용할 수 있다. 이때, 패널티는 기하학적 패널티일 수 있는데, 예를 들어, 물리적으로 부자연스러운 동작(예: 발이 떠 있음, 관절이 꺾임 등)에 대해 패널티를 부여하는 손실을 의미한다. 이에 따라, 학습 중 부자연스러운 결과를 억제하여 현실감 있는 모션 생성을 유도할 수 있다.

[0052] 또, 물리 보정기는, 역확산 과정에 통합할 수 있도록 엔드-투-엔드(End-to-End) 구조로 동작될 수 있다. 즉, 조건부 확산 모델의 역확산 과정(Reverse Diffusion Step) 중간에 물리 보정 로직을 끼워넣는다는 뜻이다. 즉, 생성된 결과를 사후 보정(Post-Process) 하는 것이 아니라, 생성 도중에 물리 보정을 반영한다는 의미이다. 이에 따라, 사람이 따로 보정하지 않아도 자연스러운 모션이 자동으로 나오게 만들게 되므로, 후처리 없이 자연스러운 모션을 직접 뽑아내는 내장형 보정형 생성기가 구현될 수 있다. 또, 엔드-투-엔드 구조로 동작시킨다는 의미는, [텍스트 → 모션 출력]까지 모든 과정이 하나의 네트워크 구조 내에서 자동으로 처리된다는 뜻이다. 이에, 물리 보정기조차도 별도 외부처리 없이 전체 학습 구조 안에 포함되어 함께 훈련 및 실행될 수 있다.

표 9

[0053]

일반적인 모션 생성 구조
[텍스트] → [모션 생성] → (결과 어색함) → [나중에 보정]
본발명의 모션 생성 구조
[텍스트] → [모션 생성 + 물리 보정 포함] → [자연스러운 최종 모션]

[0054] 출력검증부(360)는, 생성된 모션을 기 설정된 3D 모델링 포맷으로 출력하고, 모션 캡셔닝을 통하여 텍스트 및 모션 간 일치를 평가하며, 일치 여부에 따라 자동 피드백 루프를 구동시킬 수 있다. 사용자 단말(100)은, 텍스트에 포함된 모션을 수행하는 3D 객체인 휴머노이드(Humanoid)를 출력할 수 있다. ① 이때, 3D 모델링 포맷은 예를 들어, 계층적 스켈레톤 + 프레임 단위 포즈 포함인 BVH(Biovision Hierarchy)나, Autodesk 계열 포맷, 메시, 애니메이션, 머티리얼 등을 종합적으로 지원하는 FBX(Filmbox) 포맷일 수 있다. 이러한 포맷으로 출력하는 이유는, 생성된 휴머노이드 모션 데이터를 외부 툴(Blender, Unity 등)에서 재사용 가능하게 저장하기 위함이다.

[0055] 그 다음은, ② 모션-텍스트 간 일치가 되는지를 평가해야 하는데, 생성된 모션이, 입력된 텍스트와 의미적으로 일치하는지를 평가해야 한다. 예를 들어, 이 모션은 '점프한 후 착지했다' 라는 설명에 부합하는가?" 와 같은 질문에 응답을 할 수 있어야 하는데, 이를 위하여 모션 캡셔닝을 수행하는 방법도 이용할 수 있지만, CLIP 기반 텍스트-모션 매칭을 이용하여 일치 여부를 판단할 수도 있다.

표 10

[0056]

방법 1: Motion Captioning 모델 사용	방법 2: CLIP 기반 텍스트-모션 매칭
-기존 MoCap → 텍스트 생성 모델 사용 (예: ActionCLIP, TEMOS, MotionCLIP) -생성된 모션을 입력하면 자동으로 텍스트 설명을 생성 → 입력 텍스트와 비교 (BLEU, METEOR, CLIPScore 등)	-모션과 텍스트를 공통 임베딩 공간에 매핑 -두 임베딩의 코사인 유사도로 평가
caption = motion_caption_model.generate(motion_input) similarity = text_similarity(caption, original_prompt)	motion_emb = motion_encoder(motion_sequence) text_emb = clip.encode_text(prompt) score = cosine_similarity(motion_emb, text_emb)

[0057] 도 4의 (b)를 보면 모션 캡셔닝으로 도출된 텍스트와 모션 간 비교를 한다고 되어 있는데 이는 표 10의 두 가지 방법을 혼합한 방법(방법 3)이다. 물론, 상술한 방법 외에도 다양한 방법을 이용하여 텍스트가 제대로 모션으로 표현되었는지를 파악할 수 있음은 자명하다 할 것이다.

[0058] 그 다음 ③ 자동 피드백 루프를 구성해야 하는데, 방법 1의 경우 [입력 텍스트-캡션 텍스트]간 비교를, 방법 2 및 방법 3의 경우 [모션-텍스트] 간 비교를 해야 한다. 이 두 가지의 일치 평가점수가 낮거나 일치하지 않은 경우, 모션을 재생성하거나 수정을 할 수 있다. 학습(Learning)시에는 손실로 사용하고, 추론(Inference)시에는 샘플링 조절 또는 결과 걸러내기가 가능할 수 있다.

[0059] 본 발명의 일 실시예에서는 이하 표 11과 같은 종래기술에 따른 문제점을 상술한 구성으로 해결했다.

표 11

[0060]

현존하는 확산 모델 기반 텍스트-투-모션의 한계	
① 생성 가능한 동작 길이의 제한	-대부분의 모델이 수초에서 최대 10초 내외의 모션만 안정적으로 생성하며, 그 이상 길이의 동작을 생성할 경우 노이즈가 많아지고 부자연스러워짐. -이는 HumanML3D와 같은 짧은 클립 중심 데이터셋으로 훈련된 이유가 큼.
② 다중 액션·순차 동작 처리의 어려움	-“점프한 뒤 착지하여 달린다” 처럼 복합 동작을 순차적으로 묘사하는 경우, 모델은 각 부분 동작을 정확히 인식하고 순서대로 연결하는 데 실패하기 쉬움. -이를 해결하기 위해 일부 연구는 긴 문장을 하위 구문으로 분리해 개별 동작으로 학습하거나, 두 단계 추론(DoubleTake)처럼 사전 학습된 확산 모델을 활용해 경계 만 후처리하는 방식을 제시했으나, 여전히 데이터 주석이나 특수한 트릭에 의존함.
③ 물리적 자연스러움 및 현실감 부족	-생성된 모션에서 풋슬라이딩(foot sliding), 과도한 관절 회전, 불안정한 착지 등 물 리 법칙을 위반하는 현상이 잦음. -일부는 소프트 물리 제약 손실(MoFusion)이나 샘플링 단계 물리 시뮬레이션 보정 (PhysDiff)을 도입했지만, 학습 시 제약은 표현력을 제한하고, 후처리 방식은 일관 된 엔드투엔드 모델이 아님.
④ 조건 반영 및 세밀 제어의 미흡	-텍스트 의도와 일치(compliance)하지 않는 동작이 섞이거나, Classifier-Free Guidance로 다양성과 정확성 사이의 트레이드오프가 존재함. -특정 프레임·관절을 세부적으로 제어하기 어렵고, 모션 인페인팅·관절별 컨트롤 기 법도 제한적인 범위에서만 가능함.
⑤ 다인(Person-to-Person) 상호작용 및 실시간 생성 미지원	-두 명 이상의 상호작용 모션 생성은 ComMDM 같은 별도 모듈이 필요하며, 실시간 대화형 생성 기능도 제공되지 않음. -수백 단계의 역확산 과정으로 인해 샘플링 속도가 느려짐.

[0061] 이하, 상술한 도 2의 통합제어서버의 구성에 따른 동작 과정을 도 3 및 도 4를 예로 들어 상세히 설명하기로 한다. 다만, 실시예는 본 발명의 다양한 실시예 중 어느 하나일 뿐, 이에 한정되지 않음은 자명하다 할 것이다.

[0062] 도 3을 참조하면, (a) 통합제어서버(300)는 사용자 단말(100)에서 문장을 입력하면, 문장 내에서 동사를 추출하고, 동사를 행동 단위로 분할한 후, 각 행동 단위의 행동의 소요시간을 추론하여 데이터쌍을 생성한다. 그리고, (b)와 같이 통합제어서버(300)는 각 행동 단위 및 소요시간을 조건부 확산 모델에 입력하여 각 행동에 대응하는 모션을 생성한다. 모션을 생성할 때 이후 후보정을 할 필요가 없도록, 통합제어서버(300)는 (c)와 같이 텍스트 임베딩 및 모션의 잠재 표현을 결합하여 조건부 확산 모델에 입력할 수 있다. 그리고, 통합제어서버(300)는 (d) 조건부 확산 모델에서 출력된 모션을 각각 연결하여 이어붙이고, 사용자 단말(100)로 제공할 수 있다. 이때, 도 4의 (a)와 같은 물리보정을 통하여 비현실적인 관절가동을 없애 사람이 마치 좀비처럼 관절이 꺾이거나, 버그가 난 NPC처럼 행동하는 일이 없도록 한다. 통합제어서버(300)는 이렇게 완성된 결과물은 처음에 사용자 단말(100)에서 입력된 텍스트와 비교함으로써 사용자가 원하는 바를 얼마나 표현했는지를 확인하고, 일치하지 않거나 일치율이 임곗값보다 낮은 경우 모션을 다시 생성하도록 함으로써 사용자가 의도하는 바를 최대한 표현하도록 할 수 있다.

[0063] 이와 같은 도 2 내지 도 4의 모션 자동 생성 서비스 제공 방법에 대해서 설명되지 아니한 사항은 앞서 도 1을 통해 모션 자동 생성 서비스 제공 방법에 대하여 설명된 내용과 동일하거나 설명된 내용으로부터 용이하게 유추 가능하므로 이하 설명을 생략하도록 한다.

[0064] 도 5는 본 발명의 일 실시예에 따른 도 1의 계층적 확산 기반 텍스트-투-휴머노이드 모션 자동 생성 시스템에 포함된 각 구성들 상호 간에 데이터가 송수신되는 과정을 나타낸 도면이다. 이하, 도 5를 통해 각 구성들 상호 간에 데이터가 송수신되는 과정의 일 예를 설명할 것이나, 이와 같은 실시예로 본원이 한정 해석되는 것은 아니며, 앞서 설명한 다양한 실시예들에 따라 도 5에 도시된 데이터가 송수신되는 과정이 변경될 수 있음은 기술분야에 속하는 당업자에게 자명하다.

[0065] 도 5를 참조하면, 통합제어서버는, 사용자 단말로부터 텍스트를 수신하고(S5100), 텍스트에 포함된 동사(Verb)를 추출한 후, 동사를 적어도 하나의 행동 단위로 분해하고, 적어도 하나의 행동 단위의 행동의 순서 및 시간을 추론한다(S5200).

[0066] 그리고, 통합제어서버는, 적어도 하나의 행동 단위에 대응하는 모션(Motion)을 생성하도록, 기 구축된 조건부 확산 모델(Conditional Diffusion Model)을 순차적으로 호출하여 모션을 생성하고(S5300), 적어도 하나의 행동

단위에 대응하는 구간별 모션을 접합한다(S5400).

[0067] 상술한 단계들(S5100~S5400)간의 순서는 예시일 뿐, 이에 한정되지 않는다. 즉, 상술한 단계들(S5100~S5400)간의 순서는 상호 변동될 수 있으며, 이중 일부 단계들은 동시에 실행되거나 삭제될 수도 있다.

[0068] 이와 같은 도 5의 모션 자동 생성 서비스 제공 방법에 대해서 설명되지 아니한 사항은 앞서 도 1 내지 도 4를 통해 모션 자동 생성 서비스 제공 방법에 대하여 설명된 내용과 동일하거나 설명된 내용으로부터 용이하게 유추 가능하므로 이하 설명을 생략하도록 한다.

[0069] 도 5를 통해 설명된 일 실시예에 따른 모션 자동 생성 서비스 제공 방법은, 컴퓨터에 의해 실행되는 애플리케이션이나 프로그램 모듈과 같은 컴퓨터에 의해 실행가능한 명령어를 포함하는 기록 매체의 형태로도 구현될 수 있다. 컴퓨터 판독 가능 매체는 컴퓨터에 의해 액세스될 수 있는 임의의 가용 매체일 수 있고, 휘발성 및 비휘발성 매체, 분리형 및 비분리형 매체를 모두 포함한다. 또한, 컴퓨터 판독가능 매체는 컴퓨터 저장 매체를 모두 포함할 수 있다. 컴퓨터 저장 매체는 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈 또는 기타 데이터와 같은 정보의 저장을 위한 임의의 방법 또는 기술로 구현된 휘발성 및 비휘발성, 분리형 및 비분리형 매체를 모두 포함한다.

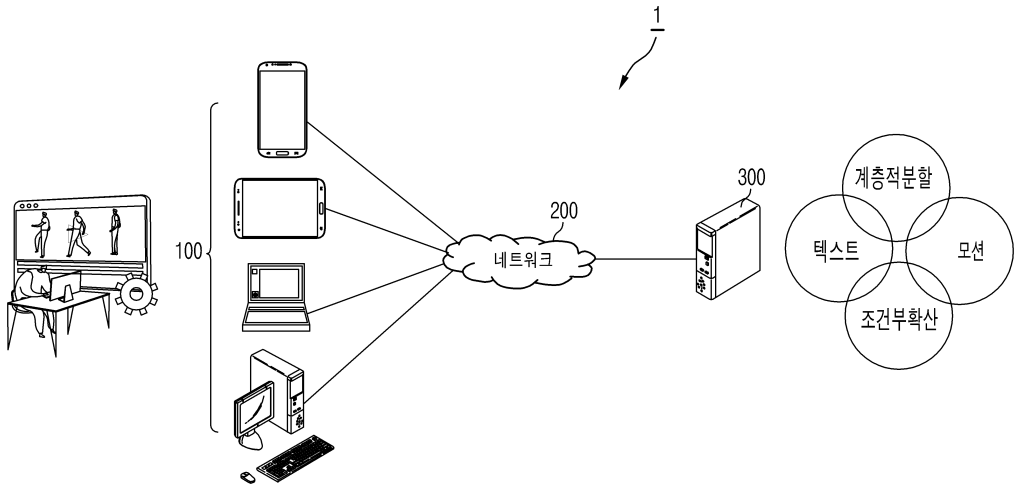
[0070] 전술한 본 발명의 일 실시예에 따른 모션 자동 생성 서비스 제공 방법은, 단말기에 기본적으로 설치된 애플리케이션(이는 단말기에 기본적으로 탑재된 플랫폼이나 운영체제 등에 포함된 프로그램을 포함할 수 있음)에 의해 실행될 수 있고, 사용자가 애플리케이션 스토어 서버, 애플리케이션 또는 해당 서비스와 관련된 웹 서버 등의 애플리케이션 제공 서버를 통해 마스터 단말기에 직접 설치한 애플리케이션(즉, 프로그램)에 의해 실행될 수도 있다. 이러한 의미에서, 전술한 본 발명의 일 실시예에 따른 모션 자동 생성 서비스 제공 방법은 단말기에 기본적으로 설치되거나 사용자에 의해 직접 설치된 애플리케이션(즉, 프로그램)으로 구현되고 단말기 등의 컴퓨터로 읽을 수 있는 기록매체에 기록될 수 있다.

[0071] 전술한 본 발명의 설명은 예시를 위한 것이며, 본 발명이 속하는 기술분야의 통상의 지식을 가진 자는 본 발명의 기술적 사상이나 필수적인 특징을 변경하지 않고서 다른 구체적인 형태로 쉽게 변형이 가능하다는 것을 이해할 수 있을 것이다. 그러므로 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적이 아닌 것으로 이해해야만 한다. 예를 들어, 단일형으로 설명되어 있는 각 구성 요소는 분산되어 실시될 수도 있으며, 마찬가지로 분산된 것으로 설명되어 있는 구성 요소들도 결합된 형태로 실시될 수 있다.

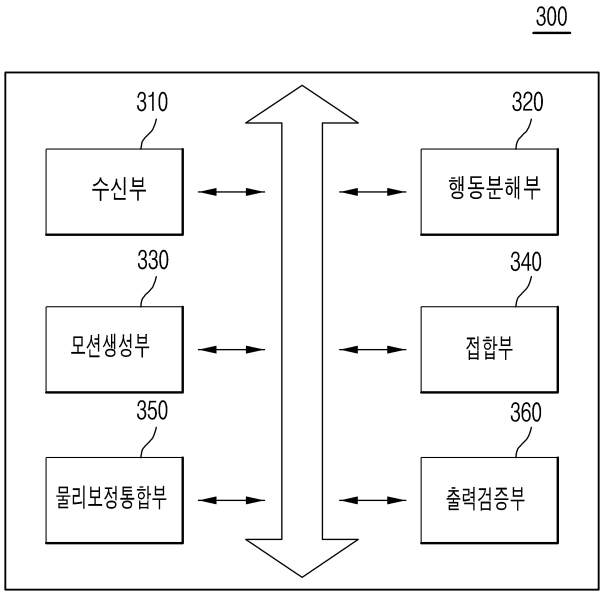
[0072] 본 발명의 범위는 상기 상세한 설명보다는 후술하는 특허청구범위에 의하여 나타내어지며, 특허청구범위의 의미 및 범위 그리고 그 균등 개념으로부터 도출되는 모든 변경 또는 변형된 형태가 본 발명의 범위에 포함되는 것으로 해석되어야 한다.

도면

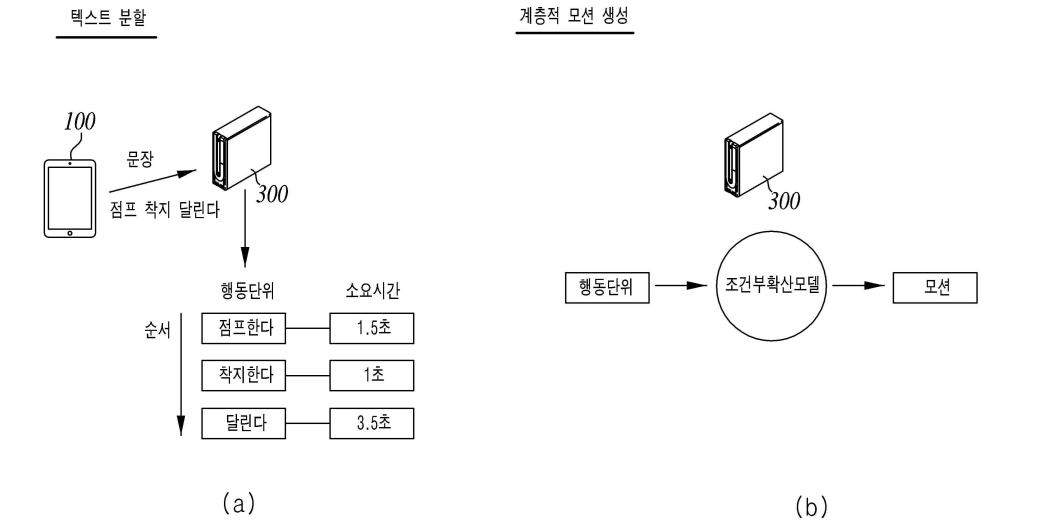
도면1



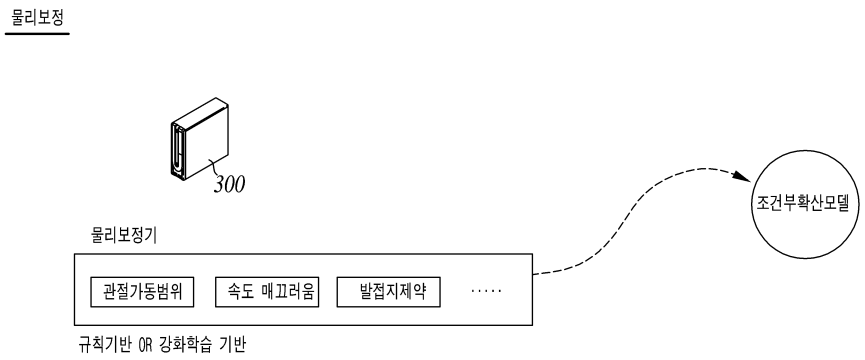
도면2



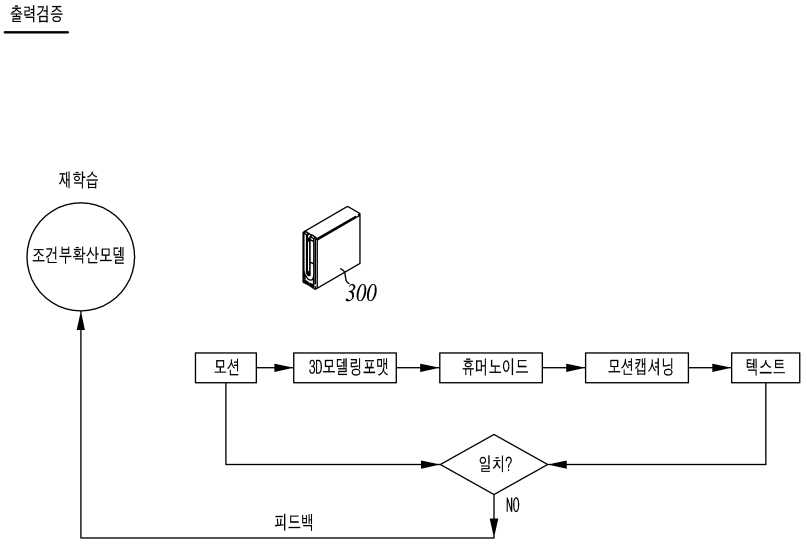
도면3


텍스트 분할
100
문장
점프 착지 달린다
300
행동단위
소요시간
순서
점프한다
1.5초
착지한다
1초
달린다
3.5초
(a)
계층적 모션 생성
300
행동단위
조건부확산모델
모션
(b)
결합
텍스트임베딩
모션의잠재표현
조건부확산모델
(c)
모션접합
500
겹침보간
겹침보간
제1구간
점프한다
제2구간
착지한다
제3구간
달린다
모션
모션
모션
(d)

도면4

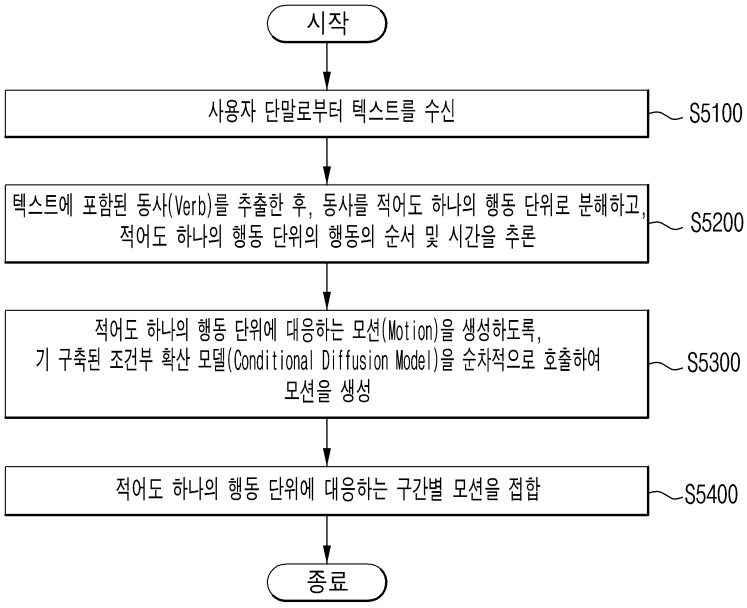

물리보정
300
조건부확산모델
물리보정기
관절가동범위
속도 매끄러움
발접지제약
.....
규칙기반 OR 강화학습 기반

(a)


출력검증
재학습
조건부확산모델
300
모션
3D모델링포맷
휴머노이드
모션캡셔닝
텍스트
일치?
NO
피드백

(b)

도면5


시작
사용자 단말로부터 텍스트를 수신
S5100
텍스트에 포함된 동사(Verb)를 추출한 후, 동사를 적어도 하나의 행동 단위로 분해하고,
적어도 하나의 행동 단위의 행동의 순서 및 시간을 추론
S5200
적어도 하나의 행동 단위에 대응하는 모션(Motion)을 생성하도록,
기 구축된 조건부 확산 모델(Conditional Diffusion Model)을 순차적으로 호출하여
모션을 생성
S5300
적어도 하나의 행동 단위에 대응하는 구간별 모션을 접합
S5400
종료