(19) 대한민국 지식재산처(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2026년07월15일
(11) 등록번호 10-2990833
(24) 등록일자 2026년07월09일

(51) 국제특허분류(Int. Cl.)
*G06T 13/40* (2011.01) *G06F 40/279* (2020.01)
*G06N 3/0475* (2023.01)
(52) CPC특허분류
*G06T 13/40* (2013.01)
*G06F 40/279* (2020.01)
(21) 출원번호 10-2025-0182856
(22) 출원일자 2025년11월26일
심사청구일자 2025년11월26일
(56) 선행기술조사문헌
KR1020230095432 A*
KR1020250127739 A*
Bhattacharya, Uttaran, et al. "Text2Gestures: A Transformer-Based Network for Generating Emotive Body Gestures for Virtual Agents." arXiv preprint arXiv:2101.11101 (2024.11.22.)*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
**주식회사 범블비**
서울특별시 관악구 행운1길 102, 2층 202호 (봉천동)
(72) 발명자
**김대호**
서울특별시 관악구 행운1길 102, 2층 202호
**김종찬**
서울특별시 관악구 행운1길 102, 2층 202호
(74) 대리인
**서평강**

전체 청구항 수 : 총 1 항 심사관 : 전세운

(54) 발명의 명칭 **텍스트 및 감정 임베딩을 이용한 컨텍스트 어웨어 모션 생성 장치 및 방법**

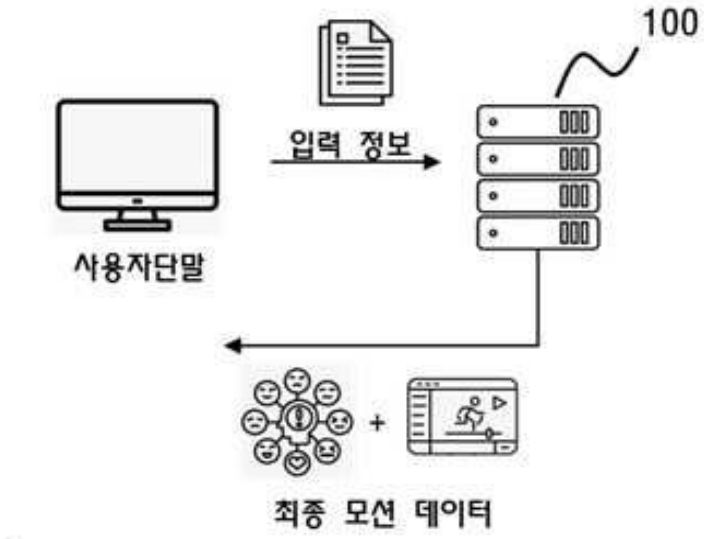
## (57) 요 약

본 발명은 텍스트 및 감정 임베딩을 이용한 컨텍스트 어웨어 모션 생성 장치 및 방법에 관한 것이다.

본 발명의 일 실시예에 따른 전자 장치는 메모리(memory) 및 상기 메모리와 연결된 프로세서(processor)를 포함하고, 상기 프로세서는, 사용자 단말로부터 3차원 모션 생성을 위하여, 캐릭터의 동작 내용을 지시하는 텍스트

(뒷면에 계속)

*대 표 도* - 도1



프롬프트(Text Prompt) 및 상기 동작에 부여될 감정 정보(Emotion Information)를 포함하는 입력 정보를 수신하고, 상기 수신된 텍스트 프롬프트와 상기 감정 정보를 각각 벡터화하여, 동작 내용에 대한 텍스트 임베딩과 동작의 뉘앙스를 결정하는 감정 임베딩을 생성하고, 생성형 인공지능(Generative AI) 모델을 이용하여, 상기 텍스트 임베딩을 기반으로 모션 시퀀스를 생성하되, 상기 감정 임베딩을 생성 과정의 조건(Condition)으로 반영하여 상기 캐릭터의 동작에 감정적 맥락(Context)을 부여하고, 상기 감정적 맥락이 부여된 상기 모션 시퀀스를 최종 모션 데이터로 생성할 수 있다.

(52) CPC특허분류
*G06N 3/0475* (2023.01)
*G06T 2207/20084* (2013.01)

이 발명을 지원한 국가연구개발사업

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

# 명 세 서

## 청구범위

### 청구항 1

전자 장치에 있어서,

메모리(memory); 및

상기 메모리와 연결된 프로세서(processor); 를 포함하고,

상기 프로세서는:

사용자 단말로부터 3차원 모션 생성을 위하여, 캐릭터의 동작 내용을 지시하는 텍스트 프롬프트(Text Prompt) 및 상기 동작에 부여될 감정 정보(Emotion Information)를 포함하는 입력 정보를 수신하고,

상기 수신된 텍스트 프롬프트와 상기 감정 정보를 각각 벡터화하여, 동작 내용에 대한 텍스트 임베딩과 동작의 뉘앙스를 결정하는 감정 임베딩을 생성하고,

생성형 인공지능(Generative AI) 모델을 이용하여, 상기 텍스트 임베딩을 기반으로 모션 시퀀스를 생성하되,

상기 감정 임베딩을 생성 과정의 조건(Condition)으로 반영하여 상기 캐릭터의 동작에 감정적 맥락(Context)을 부여하고,

상기 감정적 맥락이 부여된 상기 모션 시퀀스를 최종 모션 데이터로 생성하고,

상기 감정 정보는,

기설정된 복수의 감정 카테고리(Emotion Categories) 중 어느 하나를 지시하는 이산적인(Discrete) 감정 레이블(Emotion Label), 또는 쾌-불쾌(Valence) 및 각성(Arousal) 차원을 포함하는 연속적인 감정 벡터(Emotion Vector) 중 적어도 하나를 포함하고,

상기 프로세서는,

상기 감정 정보에 대응하는 감정 임베딩을 생성하여 상기 생성형 인공지능 모델의 확산 과정(Diffusion Process) 또는 어텐션 메커니즘(Attention Mechanism)에 주입하고,

주입된 상기 감정 임베딩을 이용하여 상기 모션 시퀀스의 모션 특성(Motion Characteristics)을 제어하고,

상기 프로세서는,

상기 입력 정보 수신 시, 상기 감정 정보의 강도(Intensity)를 나타내는 정량적 수치 값을 더 수신하고,

상기 수신된 강도에 따라 상기 감정 임베딩이 상기 생성형 인공지능 모델에 미치는 영향력을 결정하는 감정가중치(Emotion Weight)를 조절하고,

상기 확산 과정 또는 상기 어텐션 메커니즘에 상기 조절된 감정가중치를 적용하여, 동일한 종류의 상기 감정 정보에 대하여 감정 표현의 깊이(Depth)가 상이한 상기 모션 시퀀스를 생성하고,

상기 프로세서는,

상기 감정 임베딩을 이용하여 상기 캐릭터의 신체 부위별 모션을 통합적으로 생성하되,

상기 캐릭터의 얼굴 메시(Face Mesh)의 정점(Vertex) 위치를 변형하여 상기 감정 정보에 대응하는 표정(Facial Expression)을 도출하고,

상기 캐릭터의 손가락 관절의 회전값을 조절하여 상기 감정 정보에 대응하는 핸드 제스처(Hand Gesture)를 도출하고,

상기 캐릭터의 척추 및 머리 관절의 회전각을 조절하여 상기 감정 정보에 따른 무게 중심 및 시선(Gaze)을 도출하고,

상기 생성형 인공지능 모델은,

상기 텍스트 임베딩과 상기 감정 임베딩을 결합(Concatenation)하거나 크로스 어텐션(Cross-Attention)하는 레이어를 포함하는 모션 디퓨전 모델(Motion Diffusion Model, MDM)이고,

상기 프로세서는,

상기 모션 디퓨전 모델을 이용한 노이즈 제거 과정에서, 상기 감정 임베딩을 가이드 신호(Guide Signal)로 적용하고,

상기 텍스트 임베딩에 의해 결정된 기본 동작 구조(Basic Motion Structure)를 유지하면서, 상기 가이드 신호에 따라 동작의 떨림(Tremor), 속도(Speed), 및/또는 가동 범위(Range of Motion)를 포함하는 세부적인 모션 특성을 변형하여 상기 모션 시퀀스를 생성하고,

상기 프로세서는,

생성된 상기 모션 시퀀스가 상기 입력된 감정 정보의 의도와 물리적으로 일치하는지를 검증하기 위하여, 상기 모션 시퀀스 내의 각 관절의 가속도 변화량인 저크(Jerk) 값을 기반으로, 모션 격동성 지수(Motion Agitation Index)를 산출하고,

산출된 상기 모션 격동성 지수와 상기 입력 정보에 포함된 각성(Arousal) 차원의 수치 값 사이의 오차를 계산하며, 상기 오차가 기설정된 임계범위를 벗어나는 경우, 상기 오차를 줄이는 방향으로 상기 모션 시퀀스의 시간적 스케일(Temporal Scale)을 재조정하는 감정 일치도 피드백 보정을 수행하고,

상기 모션 격동성 지수는,

아래 수학식에 의하여 도출되되,

$$I_{agitation} = \frac{1}{T \cdot N} \cdot \sum_{t=1}^{T} \sum_{j=1}^{N} \left\| \frac{d^3 P_j(t)}{dt^3} \right\|$$

I_agitation은 상기 모션 격동성 지수를 의미하고, T는 상기 모션 시퀀스의 전체 프레임 수를 의미하고, N은 상기 캐릭터의 전체 관절 수를 의미하고, t는 프레임 시점을 의미하고, j는 관절 인덱스를 의미하고, P_j(t)는 t 시점에서의 j번째 관절의 3차원 위치 벡터를 의미하고,

상기 프로세서는,

상기 감정 정보가 '자신감(Confidence)' 또는 '우울(Depression)'과 같이 신체의 공간 점유와 관련된 감정 카테고리인 경우, 상기 모션 시퀀스의 공간 점유 지수(Spatial Occupancy Index)를 산출하여 자세의 수축 및 팽창 정도를 제어하고,

상기 공간 점유 지수는,

상기 캐릭터의 무게 중심(Center of Mass)으로부터 말단 관절(End-effector)들까지의 유클리드 거리(Euclidean Distance)의 평균값을 통해 도출되고,

상기 프로세서는,

상기 감정 정보가 긍정적(Positive Valence)일수록 상기 공간 점유 지수가 증가하도록, 그리고 부정적(Negative Valence)일수록 상기 공간 점유 지수가 감소하도록 상기 캐릭터의 관절 각도를 미세 보정하고,

상기 공간 점유 지수는,

아래 수학식에 의하여 도출되되,

$$I_{spatial} = \frac{1}{T \cdot M} \cdot \sum_{t=1}^{T} \sum_{k=1}^{M} \left\| P_{k(t)} - c(t) \right\|$$

I_spatial은 상기 공간 점유 지수를 의미하고, T는 상기 모션 시퀀스의 전체 프레임 수를 의미하고, M은 상기 캐릭터의 말단 관절(End-effector)의 총 개수를 의미하고, t는 프레임 시점을 의미하고, k는 말단 관절 인덱스

를 의미하고, p_k(t)는 t시점에서의 k번째 말단 관절(손, 발, 또는 머리)의 3차원 위치 벡터를 의미하고, c(t)는 t시점에서의 상기 캐릭터의 무게 중심(Center of Mass)의 3차원 위치 벡터를 의미하는 것을 특징으로 하는, 전자 장치.

**청구항 2**

삭제

**청구항 3**

삭제

## 발명의 설명

### 기 술 분 야

[0001] 본 발명은 텍스트 및 감정 임베딩을 이용한 컨텍스트 어웨어 모션 생성 장치 및 방법에 관한 것이다.

### 배 경 기 술

[0003] 본 명세서에서 달리 표시되지 않는 한, 이 섹션에 설명되는 내용들은 이 출원의 청구항들에 대한 종래 기술이 아니며, 이 섹션에 포함된다고 하여 종래 기술이라고 인정되는 것은 아니다.

[0004] 최근 메타버스(Metaverse), 가상 현실(VR), 증강 현실(AR) 등 실감형 콘텐츠 시장이 급성장함에 따라, 가상 공간 내에서 사용자를 대리하거나 상호작용하는 3차원 캐릭터(Avatar) 및 디지털 휴먼(Digital Human)의 역할이 중요해지고 있다. 이러한 디지털 휴먼이 사용자에게 깊은 몰입감을 제공하기 위해서는 단순히 사람의 외형을 모사하는 것을 넘어, 상황과 맥락에 맞는 자연스럽고 풍부한 움직임(Motion)을 구현하는 것이 필수적이다.

[0005] 종래의 3차원 캐릭터 모션 생성 기술은 전문 애니메이터가 수작업으로 키프레임(Keyframe)을 설정하거나, 모션 캡처(Motion Capture) 장비를 이용하여 배우의 연기를 녹화하는 방식이 주를 이루었다. 그러나 이러한 방식은 높은 비용과 긴 제작 시간이 소요된다는 단점이 있어, 최근에는 텍스트 프롬프트(Text Prompt)를 입력하면 인공지능이 이에 대응하는 모션을 자동으로 생성해 주는 '텍스트 기반 모션 생성(Text-to-Motion)' 기술이 활발히 연구되고 있다.

[0006] 하지만 기존의 텍스트 기반 모션 생성 모델들은 텍스트 자체의 정보에만 의존하여 모션을 생성하기 때문에, 텍스트가 담지 못하는 미묘한 감정선, 맥락(Context), 또는 동작의 뉘앙스(Nuance)를 정량적으로 제어하는 데에는 한계가 있었다. 예를 들어, "걷는다(Walking)"라는 텍스트 프롬프트에 "슬프게(Sadly)"라는 형용사를 추가하여 "슬프게 걷는다"라고 입력하더라도, 기존 모델은 학습된 데이터의 평균적인 패턴에 따라 암묵적이고 모호한 형태로만 슬픔을 표현할 뿐, 슬픔의 강도(Intensity)를 조절하거나 구체적인 비언어적 표현(시선 처리, 고개 숙임, 어깨 처짐 등)을 사용자가 의도한 대로 정밀하게 제어하기 어려웠다.

[0007] 또한, 텍스트 임베딩(Text Embedding) 내에서는 '동작의 내용(Content)'과 '감정적 스타일(Style)'이 얽혀 있어 이를 명확히 분리하거나 독립적으로 제어하는 것이 기술적으로 난해하였다. 이로 인해 생성된 결과물은 기계적이고 영혼이 없는 듯한(Robotic) 느낌을 주거나, 상황에 맞지 않는 부자연스러운 감정 표현을 보여주는 경우가 빈번하여, 높은 수준의 연기력이 요구되는 게임 시네마틱이나 영화, 버추얼 휴먼 서비스 등에 적용하기에는 부족함이 있었다.

[0008] 따라서, 사용자가 텍스트로 동작을 지시함과 동시에, 감정이나 상황적 맥락 정보를 별도의 명시적 조건(Explicit Condition)으로 입력하여, 캐릭터의 표정, 제스처, 자세 등을 통합적으로 제어함으로써 표현력이 풍부한 고품질의 모션을 생성할 수 있는 새로운 형태의 모션 생성 장치 및 방법이 절실히 요구되고 있다.

### 선행기술문헌

#### 특허문헌

[0010] (특허문헌 0001) 한국등록특허 제10-2872821호 (2025.10.14.)

## 발명의 내용

### 해결하려는 과제

[0011] 본 발명의 일 실시예는 텍스트 및 감정 임베딩을 이용한 컨텍스트 어웨어 모션 생성 장치 및 방법을 제공하는 것이다.

[0012] 본 발명에서 이루고자 하는 기술적 과제들은 이상에서 언급한 기술적 과제들로 제한되지 않으며, 언급하지 않은 또 다른 기술적 과제들은 아래의 기재로부터 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

### 과제의 해결 수단

[0014] 상술한 목적을 달성하기 위하여, 본 발명의 일 실시예에 따른 전자 장치는 메모리(memory) 및 상기 메모리와 연결된 프로세서(processor)를 포함하고, 상기 프로세서는, 사용자 단말로부터 3차원 모션 생성을 위하여, 캐릭터의 동작 내용을 지시하는 텍스트 프롬프트(Text Prompt) 및 상기 동작에 부여될 감정 정보(Emotion Information)를 포함하는 입력 정보를 수신하고, 상기 수신된 텍스트 프롬프트와 상기 감정 정보를 각각 벡터화하여, 동작 내용에 대한 텍스트 임베딩과 동작의 뉘앙스를 결정하는 감정 임베딩을 생성하고, 생성형 인공지능(Generative AI) 모델을 이용하여, 상기 텍스트 임베딩을 기반으로 모션 시퀀스를 생성하되, 상기 감정 임베딩을 생성 과정의 조건(Condition)으로 반영하여 상기 캐릭터의 동작에 감정적 맥락(Context)을 부여하고, 상기 감정적 맥락이 부여된 상기 모션 시퀀스를 최종 모션 데이터로 생성할 수 있다.

[0015] 이 때, 상기 감정 정보는, 기설정된 복수의 감정 카테고리(Emotion Categories) 중 어느 하나를 지시하는 이산적인(Discrete) 감정 레이블(Emotion Label), 또는 쾌-불쾌(Valence) 및 각성(Arousal) 차원을 포함하는 연속적인 감정 벡터(Emotion Vector) 중 적어도 하나를 포함하고, 상기 프로세서는, 상기 감정 정보에 대응하는 감정 임베딩을 생성하여 상기 생성형 인공지능 모델의 확산 과정(Diffusion Process) 또는 어텐션 메커니즘(Attention Mechanism)에 주입하고, 주입된 상기 감정 임베딩을 이용하여 상기 모션 시퀀스의 모션 특성(Motion Characteristics)을 제어할 수 있다.

[0016] 이 때, 상기 프로세서는, 상기 입력 정보 수신 시, 상기 감정 정보의 강도(Intensity)를 나타내는 정량적 수치값을 더 수신하고, 상기 수신된 강도에 따라 상기 감정 임베딩이 상기 생성형 인공지능 모델에 미치는 영향력을 결정하는 감정가중치(Emotion Weight)를 조절하고, 상기 확산 과정 또는 상기 어텐션 메커니즘에 상기 조절된 감정가중치를 적용하여, 동일한 종류의 상기 감정 정보에 대하여 감정 표현의 깊이(Depth)가 상이한 상기 모션 시퀀스를 생성할 수 있다.

[0017] 이 때, 상기 프로세서는, 상기 감정 임베딩을 이용하여 상기 캐릭터의 신체 부위별 모션을 통합적으로 생성하되, 상기 캐릭터의 얼굴 메시(Face Mesh)의 정점(Vertex) 위치를 변형하여 상기 감정 정보에 대응하는 표정(Facial Expression)을 도출하고, 상기 캐릭터의 손가락 관절의 회전값을 조절하여 상기 감정 정보에 대응하는 핸드 제스처(Hand Gesture)를 도출하고, 상기 캐릭터의 척추 및 머리 관절의 회전각을 조절하여 상기 감정 정보에 따른 무게 중심 및 시선(Gaze)을 도출할 수 있다.

[0018] 이 때, 상기 생성형 인공지능 모델은, 상기 텍스트 임베딩과 상기 감정 임베딩을 결합(Concatenation)하거나 크로스 어텐션(Cross-Attention)하는 레이어를 포함하는 모션 디퓨전 모델(Motion Diffusion Model, MDM)이고, 상기 프로세서는, 상기 모션 디퓨전 모델을 이용한 노이즈 제거 과정에서, 상기 감정 임베딩을 가이드 신호(Guide Signal)로 적용하고, 상기 텍스트 임베딩에 의해 결정된 기본 동작 구조(Basic Motion Structure)를 유지하면서, 상기 가이드 신호에 따라 동작의 떨림(Tremor), 속도(Speed), 및/또는 가동 범위(Range of Motion)를 포함하는 세부적인 모션 특성을 변형하여 상기 모션 시퀀스를 생성할 수 있다.

[0019] 이 때, 상기 프로세서는, 생성된 상기 모션 시퀀스가 상기 입력된 감정 정보의 의도와 물리적으로 일치하는지를 검증하기 위하여, 상기 모션 시퀀스 내의 각 관절의 가속도 변화량인 저크(Jerk) 값을 기반으로, 모션 격동성 지수(Motion Agitation Index)를 산출하고, 산출된 상기 모션 격동성 지수와 상기 입력 정보에 포함된 각성(Arousal) 차원의 수치 값 사이의 오차를 계산하며, 상기 오차가 기설정된 임계범위를 벗어나는 경우, 상기 오차를 줄이는 방향으로 상기 모션 시퀀스의 시간적 스케일(Temporal Scale)을 재조정하는 감정 일치도 피드백 보

정을 수행할 수 있다.

[0020] 이 때, 상기 모션 격동성 지수는, 아래 수학식에 의하여 도출되되,

[0021] $$I_{agitation} = \frac{1}{T \cdot N} \cdot \sum_{t=1}^{T} \sum_{j=1}^{N} \left\| \frac{d^3 P_j(t)}{dt^3} \right\|$$

[0022] I_agitation은 상기 모션 격동성 지수를 의미하고, T는 상기 모션 시퀀스의 전체 프레임 수를 의미하고, N은 상기 캐릭터의 전체 관절 수를 의미하고, t는 프레임 시점을 의미하고, j는 관절 인덱스를 의미하고, P_j(t)는 t 시점에서의 j번째 관절의 3차원 위치 벡터를 의미할 수 있다.

[0023] 이 때, 상기 프로세서는, 상기 감정 정보가 '자신감(Confidence)' 또는 '우울(Depression)'과 같이 신체의 공간 점유와 관련된 감정 카테고리인 경우, 상기 모션 시퀀스의 공간 점유 지수(Spatial Occupancy Index)를 산출하여 자세의 수축 및 팽창 정도를 제어할 수 있다.

[0024] 이 때, 상기 공간 점유 지수는, 상기 캐릭터의 무게 중심(Center of Mass)으로부터 말단 관절(End-effector)들까지의 유클리드 거리(Euclidean Distance)의 평균값을 통해 도출되고, 상기 프로세서는, 상기 감정 정보가 긍정적(Positive Valence)일수록 상기 공간 점유 지수가 증가하도록, 그리고 부정적(Negative Valence)일수록 상기 공간 점유 지수가 감소하도록 상기 캐릭터의 관절 각도를 미세 보정할 수 있다.

[0025] 이 때, 상기 공간 점유 지수는, 아래 수학식에 의하여 도출되되,

[0026] $$I_{spatial} = \frac{1}{T \cdot M} \cdot \sum_{t=1}^{T} \sum_{k=1}^{M} \left\| P_{k(t)} - c(t) \right\|$$

[0027] I_spatial은 상기 공간 점유 지수를 의미하고, T는 상기 모션 시퀀스의 전체 프레임 수를 의미하고, M은 상기 캐릭터의 말단 관절(End-effector)의 총 개수를 의미하고, t는 프레임 시점을 의미하고, k는 말단 관절 인덱스를 의미하고, p_k(t)는 t시점에서의 k번째 말단 관절(손, 발, 또는 머리)의 3차원 위치 벡터를 의미하고, c(t)는 t시점에서의 상기 캐릭터의 무게 중심(Center of Mass)의 3차원 위치 벡터를 의미할 수 있다.

## 발명의 효과

[0029] 이와 같이 본 발명의 일 실시예에 따르면, 텍스트 및 감정 임베딩을 이용한 컨텍스트 어웨어 모션 생성 장치 및 방법을 제공할 수 있다.

[0030] 본 발명에서 얻을 수 있는 효과는 이상에서 언급한 효과들로 제한되지 않으며, 언급하지 않은 또 다른 효과들은 아래의 기재로부터 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

## 도면의 간단한 설명

[0032] 본 발명의 특정한 바람직한 실시예들의 상기에서 설명한 바와 같은 또한 다른 측면들과, 특징들 및 이득들은 첨부 도면들과 함께 처리되는 하기의 설명으로부터 보다 명백하게 될 것이다.

도 1은 본 발명의 일 실시예에 따른 텍스트 및 감정 임베딩을 이용한 컨텍스트 어웨어 모션 생성 장치의 개념도이다.

도 2는 본 발명의 일 실시예에 따른 전자 장치의 블록도이다.

도 3은 본 발명의 일 실시예에 따른 최종 모션 데이터의 도출을 나타내는 도면이다.

도 4는 본 발명의 일 실시예에 따른 모션 격동성 지수의 개념을 나타내는 도면이다.

도 5는 본 발명의 일 실시예에 따른 말단 관절 및 무게 중심을 나타내는 도면이다.

도 6은 본 발명의 일 실시예에 따른 텍스트 및 감정 임베딩을 이용한 컨텍스트 어웨어 모션 생성 방법의 흐름도이다.

상기 도면들을 통해, 유사 참조 번호들은 동일한 혹은 유사한 엘리먼트들과, 특징들 및 구조들을 도시하기 위해

사용된다는 것에 유의해야만 한다.

## *발명을 실시하기 위한 구체적인 내용*

[0033] 이하, 본 발명의 실시예를 첨부된 도면을 참조하여 상세하게 설명한다.

[0034] 실시예를 설명함에 있어서 본 발명이 속하는 기술 분야에 익히 알려져 있고 본 발명과 직접적으로 관련이 없는 기술 내용에 대해서는 설명을 생략한다. 이는 불필요한 설명을 생략함으로써 본 발명의 요지를 흐리지 않고 더욱 명확히 전달하기 위함이다.

[0035] 마찬가지 이유로 첨부 도면에 있어서 일부 구성요소는 과장되거나 생략되거나 개략적으로 도시되었다. 또한, 각 구성요소의 크기는 실제 크기를 전적으로 반영하는 것이 아니다. 각 도면에서 동일한 또는 대응하는 구성요소에는 동일한 참조 번호를 부여하였다.

[0036] 본 발명의 이점 및 특징, 그리고 그것들을 달성하는 방법은 첨부되는 도면과 함께 상세하게 후술되어 있는 실시예들을 참조하면 명확해질 것이다. 그러나 본 발명은 이하에서 개시되는 실시 예들에 한정되는 것이 아니라 서로 다른 다양한 형태로 구현될 수 있으며, 단지 본 실시 예들은 본 발명의 개시가 완전하도록 하고, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 발명의 범주를 완전하게 알려주기 위해 제공되는 것이며, 본 발명은 청구항의 범주에 의해 정의될 뿐이다. 명세서 전체에 걸쳐 동일 참조 부호는 동일 구성 요소를 지칭한다.

[0037] 이때, 처리 흐름도 도면들의 각 블록과 흐름도 도면들의 조합들은 컴퓨터 프로그램 인스트럭션들에 의해 수행될 수 있음을 이해할 수 있을 것이다. 이들 컴퓨터 프로그램 인스트럭션들은 범용 컴퓨터, 특수용 컴퓨터 또는 기타 프로그램 가능한 데이터 프로세싱 장비의 프로세서에 탑재될 수 있으므로, 컴퓨터 또는 기타 프로그램 가능한 데이터 프로세싱 장비의 프로세서를 통해 수행되는 그 인스트럭션들이 흐름도 블록(들)에서 설명된 기능들을 수행하는 수단을 생성하게 된다. 이들 컴퓨터 프로그램 인스트럭션들은 특정 방식으로 기능을 구현하기 위해 컴퓨터 또는 기타 프로그램 가능한 데이터 프로세싱 장비를 지향할 수 있는 컴퓨터 이용 가능 또는 컴퓨터 판독 가능 메모리에 저장되는 것도 가능하므로, 그 컴퓨터 이용가능 또는 컴퓨터 판독 가능 메모리에 저장된 인스트럭션들은 흐름도 블록(들)에서 설명된 기능을 수행하는 인스트럭션 수단을 내포하는 제조 품목을 생산하는 것도 가능하다. 컴퓨터 프로그램 인스트럭션들은 컴퓨터 또는 기타 프로그램 가능한 데이터 프로세싱 장비 상에 탑재되는 것도 가능하므로, 컴퓨터 또는 기타 프로그램 가능한 데이터 프로세싱 장비 상에서 일련의 동작 단계들이 수행되어 컴퓨터로 실행되는 프로세스를 생성해서 컴퓨터 또는 기타 프로그램 가능한 데이터 프로세싱 장비를 수행하는 인스트럭션들은 흐름도 블록(들)에서 설명된 기능들을 실행하기 위한 단계들을 제공하는 것도 가능하다.

[0038] 또한, 각 블록은 특정된 논리적 기능(들)을 실행하기 위한 하나 이상의 실행 가능한 인스트럭션들을 포함하는 모듈, 세그먼트 또는 코드의 일부를 나타낼 수 있다. 또, 몇 가지 대체 실행 예들에서는 블록들에서 언급된 기능들이 순서를 벗어나서 발생하는 것도 가능함을 주목해야 한다. 예컨대, 잇달아 도시되어 있는 두 개의 블록들은 사실 실질적으로 동시에 수행되는 것도 가능하고 또는 그 블록들이 때때로 해당하는 기능에 따라 역순으로 수행되는 것도 가능하다.

[0039] 이 때, 본 실시 예에서 사용되는 '~부'라는 용어는 소프트웨어 또는 FPGA(field-Programmable Gate Array) 또는 ASIC(Application Specific Integrated Circuit)과 같은 하드웨어 구성요소를 의미하며, '~부'는 어떤 역할들을 수행한다. 그렇지만 '~부'는 소프트웨어 또는 하드웨어에 한정되는 의미는 아니다. '~부'는 어드레싱할 수 있는 저장 매체에 있도록 구성될 수도 있고 하나 또는 그 이상의 프로세서들을 재생시키도록 구성될 수도 있다. 따라서, 일 예로서 '~부'는 소프트웨어 구성요소들, 객체지향 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.

[0040] 본 발명의 실시예들을 구체적으로 설명함에 있어서, 특정 시스템의 예를 주된 대상으로 할 것이지만, 본 명세서에서 청구하고자 하는 주요한 요지는 유사한 기술적 배경을 가지는 여타의 통신 시스템 및 서비스에도 본 명세서에 개시된 범위를 크게 벗어나지 아니하는 범위에서 적용 가능하며, 이는 당해 기술분야에서 숙련된 기술적

지식을 가진 자의 판단으로 가능할 것이다.

[0041] 이 때, 후술하는 사용자단말은 통신 가능한 데스크탑 컴퓨터(desktop computer), 랩탑 컴퓨터(laptop computer), 노트북(notebook), 스마트폰(smart phone), 태블릿 PC(tablet PC), 모바일폰(mobile phone), 스마트 워치(smart watch), 스마트 글래스(smart glass), e-book 리더기, PMP(portable multimedia player), 휴대용 게임기, 네비게이션(navigation) 장치, 디지털 카메라(digital camera), DMB(digital multimedia broadcasting) 재생기, 디지털 음성 녹음기(digital audio recorder), 디지털 음성 재생기(digital audio player), 디지털 동영상 녹화기(digital video recorder), 디지털 동영상 재생기(digital video player), PDA(Personal Digital Assistant) 등을 포함할 수 있다.

[0043] 도 1은 본 발명의 일 실시예에 따른 텍스트 및 감정 임베딩을 이용한 컨텍스트 어웨어 모션 생성 장치의 개념도이고, 도 2는 본 발명의 일 실시예에 따른 전자 장치(100)의 블록도이다.

[0044] 일 실시예에 따른 전자 장치(100)는 프로세서(110) 및 메모리(120)를 포함한다. 프로세서(110)는 전술한 적어도 하나의 방법을 수행할 수 있다. 메모리(120)는 상술한 방법과 관련된 정보를 저장하거나 상술한 방법이 구현된 프로그램을 저장할 수 있다. 메모리(120)는 휘발성 메모리 또는 비휘발성 메모리일 수 있다. 메모리(120)는 '데이터베이스', '저장부' 등으로 호칭될 수 있다.

[0045] 프로세서(110)는 프로그램을 실행하고, 전자 장치(100)를 제어할 수 있다. 프로세서(110)에 의하여 실행되는 프로그램의 코드는 메모리(120)에 저장될 수 있다. 장치(100)는 입출력 장치(미도시)를 통하여 외부 장치(예를 들어, 퍼스널 컴퓨터 또는 네트워크)에 연결되고, 데이터를 교환할 수 있다.

[0046] 이 때, 상기 프로세서(110)는, 사용자 단말로부터 3차원 모션 생성을 위하여, 캐릭터의 동작 내용을 지시하는 텍스트 프롬프트(Text Prompt) 및 상기 동작에 부여될 감정 정보(Emotion Information)를 포함하는 입력 정보를 수신할 수 있다.

[0047] 이 때, 상기 감정 정보는, 기설정된 복수의 감정 카테고리(Emotion Categories) 중 어느 하나를 지시하는 이산적인(Discrete) 감정 레이블(Emotion Label), 또는 쾌-불쾌(Valence) 및 각성(Arousal) 차원을 포함하는 연속적인 감정 벡터(Emotion Vector) 중 적어도 하나를 포함할 수 있다.

[0048] 구체적으로, 상기 텍스트 프롬프트는 사용자가 자연어(Natural Language) 형태로 입력하는 동작 지시어로서, 생성하고자 하는 모션의 기본적인 형태, 행위의 종류, 또는 시맨틱(Semantic) 내용을 결정하는 데이터를 의미한다. 예를 들어, "앞으로 걷는다", "제자리에서 점프한다", "상대방에게 손을 흔들어 인사한다"와 같은 구체적인 행동 묘사가 이에 해당할 수 있다.

[0049] 또한, 상기 감정 정보는 상기 텍스트 프롬프트가 지시하는 기본 동작에 부여될 분위기, 스타일, 또는 심리적 상태를 결정하는 보조 데이터를 의미한다. 여기서, 상기 이산적인 감정 레이블은 인간이 보편적으로 인지하는 감정의 카테고리를 텍스트나 인덱스(Index) 형태로 지정한 것으로서, 예를 들어 '기쁨', '슬픔', '분노', '공포', '중립' 등으로 명확하게 구분되는 감정 분류를 포함할 수 있다.

[0050] 반면, 상기 연속적인 감정 벡터는 감정을 다차원 공간상의 좌표로 수치화하여 표현한 데이터일 수 있다. 예를 들어, 상기 쾌-불쾌(Valence) 차원은 해당 감정이 긍정적인지(Positive, 예: 행복, 만족) 또는 부정적인지(Negative, 예: 우울, 불만)를 나타내는 척도를 의미하며, 상기 각성(Arousal) 차원은 해당 감정의 에너지 레벨이나 신체적 활성도가 높은지(High, 예: 흥분, 긴장) 또는 낮은지(Low, 예: 이완, 나른함)를 나타내는 척도를 의미한다. 이를 통해, 사용자는 단순한 '기쁨'이라는 단어를 넘어, '에너지가 넘치는 격정적인 기쁨(High Arousal & Positive Valence)'과 '편안하고 차분한 기쁨(Low Arousal & Positive Valence)'과 같이 미세한 감정의 뉘앙스를 정량적으로 구분하여 입력할 수 있다.

[0052] 또한, 도 3을 참조하면, 상기 프로세서는, 상기 수신된 텍스트 프롬프트와 상기 감정 정보를 각각 벡터화하여, 동작 내용에 대한 텍스트 임베딩과 동작의 뉘앙스를 결정하는 감정 임베딩을 생성하고, 생성형 인공지능(Generative AI) 모델을 이용하여, 상기 텍스트 임베딩을 기반으로 모션 시퀀스를 생성하되, 상기 감정 임베딩을 생성 과정의 조건(Condition)으로 반영하여 상기 캐릭터의 동작에 감정적 맥락(Context)을 부여하고, 상기 감정적 맥락이 부여된 상기 모션 시퀀스를 최종 모션 데이터로 생성할 수 있다.

[0053] 보다 상세하게 살펴보면, 상기 프로세서는, 상기 감정 정보에 대응하는 감정 임베딩을 생성하여 상기 생성형 인공지능 모델의 확산 과정(Diffusion Process) 또는 어텐션 메커니즘(Attention Mechanism)에 주입하고, 주입된 상기 감정 임베딩을 이용하여 상기 모션 시퀀스의 모션 특성(Motion Characteristics)을 제어할 수 있다.

[0054] 구체적으로, 상기 텍스트 임베딩은 자연어 처리 모델(예: CLIP, BERT 등)을 통해 변환된 고차원 벡터로서, 모션의 '구조적 뼈대(Skeleton Structure)'나 '행위의 종류(Semantic Action)'를 결정하는 기준 정보로 작용한다. 반면, 상기 감정 임베딩은 동작의 '스타일(Style)'이나 '분위기(Atmosphere)'를 결정하는 벡터로서, 상기 텍스트 임베딩이 형성한 뼈대 위에 입혀지는 뉘앙스 정보로 기능한다.

[0055] 이 때, 상기 프로세서가 감정 임베딩을 확산 과정(Diffusion Process)에 주입한다는 것은, 무작위 노이즈로부터 유의미한 동작 데이터를 복원하는 매 디노이징 단계(Denoising Step)마다 상기 감정 벡터를 조건(Condition)으로 연산에 참여시킨다는 것을 의미한다.

[0056] 또한, 어텐션 메커니즘(Attention Mechanism), 특히 크로스 어텐션(Cross-Attention)을 활용할 경우, 생성 모델은 각 프레임의 관절 위치를 결정할 때 상기 텍스트 임베딩뿐만 아니라 상기 감정 임베딩을 참조하여, 동작의 어느 부분에 감정을 집중적으로 반영할지 가중치를 동적으로 결정하게 된다.

[0057] 이에 따라 제어되는 상기 모션 특성(Motion Characteristics)은 동작의 물리적인 속성, 예를 들어 동작의 속도(Speed), 보폭(Stride), 관절의 가동 범위(Amplitude), 또는 신체의 미세한 떨림(Jitter) 등을 포괄할 수 있다.

[0058] 예를 들어, 사용자가 '걷는다'라는 텍스트와 '슬픔(Sadness)'이라는 감정 정보를 입력한 경우, 상기 프로세서는 걷는 행위 자체는 유지하되(텍스트 임베딩), 상기 감정 임베딩의 영향으로 인해 보폭을 좁히고, 상체의 숙임 각도를 크게 하며, 동작의 속도를 느리게 변형한 최종 모션 데이터를 생성할 수 있다.

[0059] 반대로 '기쁨(Joy)'이라는 감정 정보가 입력된 경우에는, 동일한 '걷기' 동작이라도 팔을 크게 흔들거나(가동 범위 증가), 리듬감 있게 도약하는 형태(속도 및 탄성 증가)로 모션 특성이 제어될 수 있다.

[0061] 또한, 상기 프로세서는, 상기 입력 정보 수신 시, 상기 감정 정보의 강도(Intensity)를 나타내는 정량적 수치 값을 더 수신하고, 상기 수신된 강도에 따라 상기 감정 임베딩이 상기 생성형 인공지능 모델에 미치는 영향력을 결정하는 감정가중치(Emotion Weight)를 조절하고, 상기 확산 과정 또는 상기 어텐션 메커니즘에 상기 조절된 감정가중치를 적용하여, 동일한 종류의 상기 감정 정보에 대하여 감정 표현의 깊이(Depth)가 상이한 상기 모션 시퀀스를 생성할 수 있다.

[0062] 구체적으로, 상기 강도(Intensity)는 사용자가 의도하는 감정 표현의 세기를 정량화한 파라미터(예: 0.0 ~ 1.0 사이의 실수 값)를 의미한다.

[0063] 이 때, 상기 프로세서는 이 강도 값이 높을수록 상기 감정가중치(Emotion Weight)를 상향 조정하여, 생성형 인공지능 모델의 연산 과정에서 텍스트 정보 대비 감정 정보가 미치는 영향력을 강화할 수 있다.

[0064] 예를 들어, 확산 모델의 경우, 잡음(Noise)을 제거하여 이미지를 생성하는 매 단계마다 적용되는 '조건부 안내 스케일(Guidance Scale)'에 상기 감정가중치를 반영함으로써, 결과물이 감정 조건에 더욱 강하게 부합하도록 유도할 수 있다.

[0065] 이에 따라 조절되는 감정 표현의 깊이(Depth)는 동일한 감정 카테고리 내에서의 동작의 과장 정도(Exaggeration)나 행위의 구체성을 의미할 수 있다.

[0066] 예를 들어, '슬픔(Sadness)'이라는 동일한 감정 정보가 입력되더라도, 강도가 낮게(예: 0.2) 설정된 경우 상기 프로세서는 캐릭터의 고개를 약간 숙이거나 걷는 속도를 늦추는 정도의 정적인 움직임(소극적 표현)을 생성할 수 있다.

[0067] 반면, 강도가 높게(예: 0.9) 설정된 경우, 상기 프로세서는 캐릭터가 얼굴을 감싸 쥐거나, 비틀거리며 바닥에 주저앉는 등의 동적이고 격렬한 움직임(적극적 표현)을 생성할 수 있다. 이를 통해 사용자는 단순한 감정의 분류를 넘어, 상황에 적합한 연기 톤(Acting Tone)을 정밀하게 제어할 수 있다.

[0069] 또한, 상기 프로세서는, 상기 감정 임베딩을 이용하여 상기 캐릭터의 신체 부위별 모션을 통합적으로 생성하되, 상기 캐릭터의 얼굴 메시(Face Mesh)의 정점(Vertex) 위치를 변형하여 상기 감정 정보에 대응하는 표정(Facial Expression)을 도출하고, 상기 캐릭터의 손가락 관절의 회전값을 조절하여 상기 감정 정보에 대응하는 핸드 제스처(Hand Gesture)를 도출하고, 상기 캐릭터의 척추 및 머리 관절의 회전각을 조절하여 상기 감정 정보에 따른 무게 중심 및 시선(Gaze)을 도출할 수 있다.

[0070] 구체적으로, 이는 인간의 감정 표현이 단순히 얼굴 표정에만 국한되지 않고, 손짓, 자세, 시선 등 신체 전반의 움직임이 복합적으로 작용하여 나타난다는 점에 착안한 구성이다. 상기 프로세서는 상기 감정 임베딩을 전역적

인(Global) 제어 신호로 활용하되, 이를 각 신체 부위의 특성에 맞는 국소적인(Local) 제어 파라미터로 변환하여 적용한다.

[0071] 우선, 상기 얼굴 메시의 정점 위치 변형은 3차원 모델의 기하학적 구조를 직접 제어하는 것으로서, 예를 들어 '기쁨' 감정 입력 시 입꼬리 주변의 정점들을 상방으로 이동시키고 눈가 주변의 정점들을 수축시켜 미소를 구현하거나, '분노' 감정 입력 시 미간의 정점들을 중심으로 모으고 눈썹의 정점들을 하방으로 이동시켜 찌푸린 표정을 도출하는 방식을 포함한다. 이는 텍스처 매핑만으로는 구현하기 힘든 입체적인 표정 변화를 가능하게 한다.

[0072] 또한, 핸드 제스처의 경우, 상기 프로세서는 손가락의 다관절(Multi-joint) 회전값을 세밀하게 조절하여 감정의 디테일을 완성한다. 예컨대, '불안(Anxiety)' 감정이 입력된 경우, 상기 프로세서는 손가락 관절들을 불규칙하게 굽혔다 펴는 반복적인 움직임(Fidgeting)을 생성하거나, '분노'의 경우 모든 손가락 관절을 안쪽으로 강하게 회전시켜 주먹을 쥐는 동작을 생성할 수 있다.

[0073] 나아가, 척추 및 머리 관절의 제어는 캐릭터의 전체적인 실루엣과 비언어적 의사소통을 결정짓는다. 상기 프로세서는 '자신감(Confidence)'이나 '행복'과 같은 긍정적 감정에 대해서는 척추 관절을 곧게 펴고 머리 관절을 들어 올려 시선이 정면이나 상방을 향하도록 무게 중심을 제어한다. 반면, '우울'이나 '좌절'과 같은 부정적 감정에 대해서는 척추를 둥글게 굽히고 머리를 떨구어 시선이 바닥을 향하도록 함으로써, 감정 상태에 따른 신체의 위축된 자세를 사실적으로 구현한다. 이러한 부위별 제어는 개별적으로 작동하는 것이 아니라 상기 감정 임베딩에 의해 시간적으로 동기화(Synchronization)되어 수행되므로, 결과적으로 통일성 있는 감정 연기가 가능하다.

[0075] 또한, 상기 생성형 인공지능 모델은, 상기 텍스트 임베딩과 상기 감정 임베딩을 결합(Concatenation)하거나 크로스 어텐션(Cross-Attention)하는 레이어를 포함하는 모션 디퓨전 모델(Motion Diffusion Model, MDM)일 수 있다.

[0076] 또한, 상기 프로세서는, 상기 모션 디퓨전 모델을 이용한 노이즈 제거 과정에서, 상기 감정 임베딩을 가이드 신호(Guide Signal)로 적용하고, 상기 텍스트 임베딩에 의해 결정된 기본 동작 구조(Basic Motion Structure)를 유지하면서, 상기 가이드 신호에 따라 동작의 떨림(Tremor), 속도(Speed), 및/또는 가동 범위(Range of Motion)를 포함하는 세부적인 모션 특성을 변형하여 상기 모션 시퀀스를 생성할 수 있다.

[0077] 구체적으로, 상기 모션 디퓨전 모델은 동작 생성의 기준이 되는 텍스트 임베딩과 동작의 스타일을 결정하는 감정 임베딩을 융합하여 처리하도록 설계될 수 있다.

[0078] 여기서 결합(Concatenation) 방식은 두 벡터를 물리적으로 이어 붙여 하나의 긴 입력 벡터로 모델에 주입하는 방식을 의미하며, 이는 구조가 단순하면서도 두 정보가 동시에 고려되도록 하는 효과가 있다.

[0079] 반면, 크로스 어텐션(Cross-Attention) 방식은 모델이 현재 생성 중인 동작 프레임과 연관성이 높은 감정 정보에 가중치를 더 크게 부여하여 동적으로 참조하는 방식이다. 이를 통해, 단순히 정보를 섞는 것을 넘어, 동작의 특정 구간(예: 손을 뻗는 순간)에서 감정의 영향력을 극대화하는 등 정교한 제어가 가능해진다.

[0080] 또한, 상기 가이드 신호(Guide Signal)는 노이즈로부터 모션을 복원하는 역확산(Reverse Diffusion) 과정에서, 생성되는 결과물이 감정 조건에서 벗어나지 않도록 방향성을 제시하는 벡터장(Vector Field)의 역할을 수행할 수 있다.

[0081] 이 때, 상기 프로세서는 텍스트 임베딩을 통해 '걷기', '달리기'와 같은 기본 동작 구조(Basic Motion Structure), 즉 전체적인 골격의 이동 궤적과 관절의 시퀀스 순서를 우선적으로 확립할 수 있다. 그 후, 상기 가이드 신호를 적용하여 해당 궤적에 세부적인 모션 특성을 덧입히게 된다.

[0082] 예를 들어, '공포' 감정이 가이드 신호로 입력된 경우, 상기 프로세서는 걷는 동작의 궤적에 고주파의 미세한 진동을 추가하여 떨림(Tremor)을 구현하거나, 프레임 간의 간격을 조절하여 동작의 속도(Speed)를 급격하게 변화시킬 수 있다. 또는 '위축'된 감정의 경우, 관절이 펴지는 최대 각도를 제한하여 가동 범위(Range of Motion)를 축소시킴으로써 소극적인 움직임을 생성한다. 즉, 본 발명은 동작의 '내용(Content)'을 훼손하지 않으면서 '표현 방식(Style)'만을 독립적으로 변형하여, 목적에 부합하면서도 감정적 깊이가 있는 고품질의 모션을 생성할 수 있다.

[0084] 도 4는 본 발명의 일 실시예에 따른 모션 격동성 지수의 개념을 나타내는 도면이다.

[0085] 도 4를 참조하면, 상기 프로세서는, 생성된 상기 모션 시퀀스가 상기 입력된 감정 정보의 의도와 물리적으로 일치하는지를 검증하기 위하여, 상기 모션 시퀀스 내의 각 관절의 가속도 변화량인 저크(Jerk) 값을 기반으로, 모션 격동성 지수(Motion Agitation Index)를 산출하고, 산출된 상기 모션 격동성 지수와 상기 입력 정보에 포함된 각성(Arousal) 차원의 수치 값 사이의 오차를 계산하며, 상기 오차가 기설정된 임계범위를 벗어나는 경우, 상기 오차를 줄이는 방향으로 상기 모션 시퀀스의 시간적 스케일(Temporal Scale)을 재조정하는 감정 일치도 피드백 보정을 수행할 수 있다.

[0086] 보다 구체적으로 살펴보면, 상기 모션 격동성 지수는, 아래 수학식 1에 의하여 도출될 수 있다.

[0087] [수학식 1]

[0088] $$I_{agitation} = \frac{1}{T \cdot N} \cdot \sum_{t=1}^{T} \sum_{j=1}^{N} \left\| \frac{d^3 P_j(t)}{dt^3} \right\|$$

[0089] 이 때, I_agitation은 상기 모션 격동성 지수를 의미하고, T는 상기 모션 시퀀스의 전체 프레임 수를 의미하고, N은 상기 캐릭터의 전체 관절 수를 의미하고, t는 프레임 시점을 의미하고, j는 관절 인덱스를 의미하고, P_j(t)는 t시점에서의 j번째 관절의 3차원 위치 벡터를 의미할 수 있다.

[0090] 구체적으로, 상기 구성은 생성된 결과물이 단순히 시각적으로만 유사한 것을 넘어, 사용자가 의도한 감정의 에너지 레벨을 물리적으로 정확하게 반영하고 있는지를 정량적으로 검증하기 위한 것이다.

[0091] 일반적으로 속도나 가속도는 동작의 빠르기나 힘을 나타내지만, 위치 벡터의 시간당 변화율을 세 번 미분한 값인 저크(Jerk)는 가속도의 돌발적인 변화량, 즉 동작의 덜컹거림이나 급격한 떨림을 대변하는 물리량을 의미할 수 있다.

[0092] 따라서, 이 값은 평온하거나 우울한 감정 상태에서의 부드러운 움직임과, 분노하거나 놀란 감정 상태에서의 거칠고 불안정한 움직임을 구분하는 핵심적인 지표로 활용될 수 있다.

[0093] 또한, 상기 수학식 1의 설계 구조를 살펴보면, 해당 수식은 특정 관절이나 특정 시점에 국한되지 않고, 캐릭터 전체의 움직임 특성을 포괄하도록 설계되었다.

[0094] 우선, 각 관절의 위치 벡터에 대해 3차 미분 연산을 수행하여 저크 벡터를 구하고, 그 벡터의 크기(Norm)를 산출함으로써 움직임의 방향성과 무관하게 순수한 변화의 격렬함만을 추출한다.

[0095] 이후, 전신에 분포된 모든 관절과 모션이 재생되는 전체 시간 프레임에 대하여 해당 값들을 모두 합산하고, 이를 총 관절 수와 총 프레임 수의 곱으로 나누어 정규화(Normalization) 과정을 거친다.

[0096] 이는 특정 순간의 노이즈나 특정 부위의 과도한 움직임이 전체 지수를 왜곡하는 것을 방지하고, 모션 시퀀스 전체를 관통하는 평균적인 격동성을 하나의 스칼라 값으로 도출하기 위함이다.

[0097] 결과적으로, 상기 프로세서는 이렇게 도출된 물리적 지수와 사용자가 입력한 각성(Arousal) 수치를 비교함으로써, 감정 표현의 과소 또는 과잉 여부를 판단할 수 있다.

[0098] 만약 산출된 격동성 지수가 목표 각성 수치보다 현저히 낮다면, 상기 프로세서는 모션의 시간적 스케일을 축소하여 동작을 더 빠르고 날카롭게 보정하고, 반대로 지수가 너무 높다면 시간적 스케일을 확장하여 동작을 진정시키는 방향으로 피드백을 수행할 수 있다.

[0099] 이러한 검증 및 보정 루프를 통해, 본 발명은 사용자의 추상적인 감정 의도를 구체적인 물리 법칙에 부합하는 고품질의 애니메이션으로 자동 최적화할 수 있는 장점이 있다.

[0101] 도 5는 본 발명의 일 실시예에 따른 말단 관절 및 무게 중심을 나타내는 도면이다.

[0102] 도 5를 참조하면, 상기 프로세서는, 상기 감정 정보가 '자신감(Confidence)' 또는 '우울(Depression)'과 같이 신체의 공간 점유와 관련된 감정 카테고리인 경우, 상기 모션 시퀀스의 공간 점유 지수(Spatial Occupancy Index)를 산출하여 자세의 수축 및 팽창 정도를 제어할 수 있다.

[0103] 이 때, 상기 공간 점유 지수는, 상기 캐릭터의 무게 중심(Center of Mass)으로부터 말단 관절(End-effector)들까지의 유클리드 거리(Euclidean Distance)의 평균값을 통해 도출될 수 있다.

[0104] 또한, 상기 프로세서는, 상기 감정 정보가 긍정적(Positive Valence)일수록 상기 공간 점유 지수가 증가하도록,

그리고 부정적(Negative Valence)일수록 상기 공간 점유 지수가 감소하도록 상기 캐릭터의 관절 각도를 미세 보정할 수 있다.

[0105] 보다 상세하게 살펴보면, 상기 공간 점유 지수는, 아래 수학식 2에 의하여 도출될 수 있다.

[0106] [수학식 2]

[0107] $$I_{spatial} = \frac{1}{T \cdot M} \cdot \sum_{t=1}^{T} \sum_{k=1}^{M} \left\| P_{k(t)} - c(t) \right\|$$

[0108] 이 때, I_spatial은 상기 공간 점유 지수를 의미하고, T는 상기 모션 시퀀스의 전체 프레임 수를 의미하고, M은 상기 캐릭터의 말단 관절(End-effector)의 총 개수를 의미하고, t는 프레임 시점을 의미하고, k는 말단 관절 인덱스를 의미하고, p_k(t)는 t시점에서의 k번째 말단 관절(손, 발, 또는 머리)의 3차원 위치 벡터를 의미하고, c(t)는 t시점에서의 상기 캐릭터의 무게 중심(Center of Mass)의 3차원 위치 벡터를 의미할 수 있다.

[0109] 구체적으로, 상기 구성은 심리학적으로 잘 알려진 신체의 개방성(Openness) 및 폐쇄성(Closeness)과 감정 간의 상관관계를 정량적인 수치로 변환하여 모션 생성에 적용하기 위한 것이다.

[0110] 일반적으로 인간은 자신감이 넘치거나 행복한 긍정적 감정 상태일 때 팔다리를 넓게 펴고 가슴을 여는 등 신체가 차지하는 공간을 확장하려는 경향(팽창)이 있는 반면, 우울하거나 두려운 부정적 감정 상태일 때는 몸을 웅크리고 팔다리를 몸통 쪽으로 모으는 등 공간 점유를 축소하려는 경향(수축)이 있다.

[0111] 따라서, 상기 공간 점유 지수는 이러한 자세의 기하학적 특성을 분석하여, 생성된 모션이 캐릭터의 내면 심리를 시각적으로 올바르게 대변하고 있는지를 판단하고 제어하는 핵심적인 척도로 활용된다.

[0112] 이 때, 상기 수학식 2의 설계 구조를 살펴보면, 해당 수식은 캐릭터의 신체 중심으로부터 각 말단 부위가 얼마나 멀리 뻗어 나가는지를 종합적으로 측정하도록 고안되었다.

[0113] 우선, 매 프레임마다 캐릭터의 무게 중심(예: 골반 또는 배꼽 위치) 좌표와 손, 발, 머리와 같은 말단 관절들의 좌표 간 유클리드 거리를 각각 계산한다. 이때, 무게 중심을 기준점으로 삼음으로써 캐릭터가 월드 좌표상 어디에 위치하든 상관없이 순수한 자세의 퍼짐 정도만을 추출할 수 있다.

[0114] 이후, 모든 말단 관절에 대해 계산된 거리들을 합산하고, 이를 전체 프레임 수와 관절 개수로 나누어 평균화함으로써, 특정 순간의 일시적인 동작보다는 모션 시퀀스 전반에 걸친 평균적인 자세의 크기(Scale)를 하나의 지표로 도출한다.

[0115] 결과적으로, 상기 프로세서는 감정 정보(긍정/부정)에 따라 목표로 하는 공간 점유 지수를 설정하고, 생성된 모션의 실제 지수가 이에 미치지 못하거나 초과할 경우 관절 각도를 미세하게 보정할 수 있다.

[0116] 예컨대, '자신감' 입력 시 공간 점유 지수가 낮다면 어깨를 펴고 팔을 벌리는 방향으로 관절을 회전시키고, '우울' 입력 시 지수가 높다면 고개를 숙이고 어깨를 안으로 굽히는 방향으로 보정한다.

[0117] 이를 통해, 텍스트 프롬프트만으로는 명시하기 어려운 미묘한 자세의 뉘앙스를 자동으로 최적화하여 캐릭터 연기의 진정성과 몰입감을 극대화할 수 있는 장점이 있다.

[0119] 도 6은 본 발명의 일 실시예에 따른 텍스트 및 감정 임베딩을 이용한 컨텍스트 어웨어 모션 생성 방법의 흐름도이다.

[0120] 도 6을 참조하면, 본 발명의 일 실시예에 따른 텍스트 및 감정 임베딩을 이용한 컨텍스트 어웨어 모션 생성 방법은 사용자 단말로부터 3차원 모션 생성을 위하여, 캐릭터의 동작 내용을 지시하는 텍스트 프롬프트(Text Prompt) 및 상기 동작에 부여될 감정 정보(Emotion Information)를 포함하는 입력 정보를 수신할 수 있다(S101).

[0121] 또한, 본 발명의 일 실시예에 따른 텍스트 및 감정 임베딩을 이용한 컨텍스트 어웨어 모션 생성 방법은 상기 수신된 텍스트 프롬프트와 상기 감정 정보를 각각 벡터화하여, 동작 내용에 대한 텍스트 임베딩과 동작의 뉘앙스를 결정하는 감정 임베딩을 생성할 수 있다(S103).

[0122] 또한, 본 발명의 일 실시예에 따른 텍스트 및 감정 임베딩을 이용한 컨텍스트 어웨어 모션 생성 방법은 생성형 인공지능(Generative AI) 모델을 이용하여, 상기 텍스트 임베딩을 기반으로 모션 시퀀스를 생성할 수 있다

(S105).

[0123] 또한, 본 발명의 일 실시예에 따른 텍스트 및 감정 임베딩을 이용한 컨텍스트 어웨어 모션 생성 방법은 상기 감정 임베딩을 생성 과정의 조건(Condition)으로 반영하여 상기 캐릭터의 동작에 감정적 맥락(Context)을 부여할 수 있다(S107).

[0124] 또한, 본 발명의 일 실시예에 따른 텍스트 및 감정 임베딩을 이용한 컨텍스트 어웨어 모션 생성 방법은 상기 감정적 맥락이 부여된 상기 모션 시퀀스를 최종 모션 데이터로 생성할 수 있다(S109).

[0125] 또한, 본 발명의 일 실시예에 따른 텍스트 및 감정 임베딩을 이용한 컨텍스트 어웨어 모션 생성 방법은 도 1 내지 도 5에 개시된 텍스트 및 감정 임베딩을 이용한 컨텍스트 어웨어 모션 생성 장치와 동일하게 구성될 수 있다.

[0127] 이상에서 설명된 실시예들은 하드웨어 구성요소, 소프트웨어 구성요소, 및/또는 하드웨어 구성요소 및 소프트웨어 구성요소의 조합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치, 방법 및 구성요소는, 예를 들어, 프로세서, 컨트롤러, ALU(arithmetic logic unit), 디지털 신호 프로세서(digital signal processor), 마이크로컴퓨터, FPGA(field programmable gate 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.

[0128] 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. 상기된 하드웨어 장치는 실시예의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

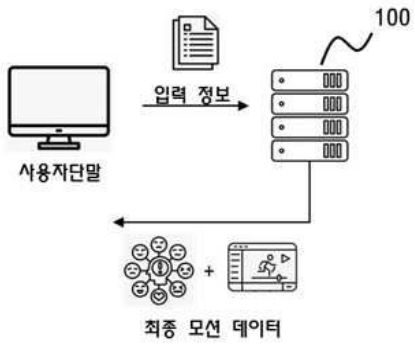
[0129] 소프트웨어는 컴퓨터 프로그램(computer program), 코드(code), 명령(instruction), 또는 이들 중 하나 이상의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로(collectively) 처리 장치를 명령할 수 있다. 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성요소(component), 물리적 장치, 가상 장치(virtual equipment), 컴퓨터 저장 매체 또는 장치, 또는 전송되는 신호 파(signal wave)에 영구적으로, 또는 일시적으로 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.

[0130] 이상과 같이 실시예들이 비록 한정된 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 상기를 기초로 다양한 기술적 수정 및 변형을 적용할 수 있다. 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.

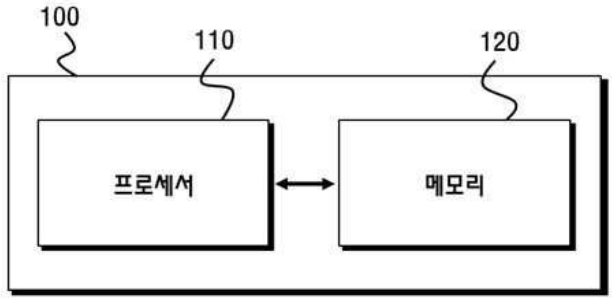
[0131] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 청구범위의 범위에 속한다.

## 도면

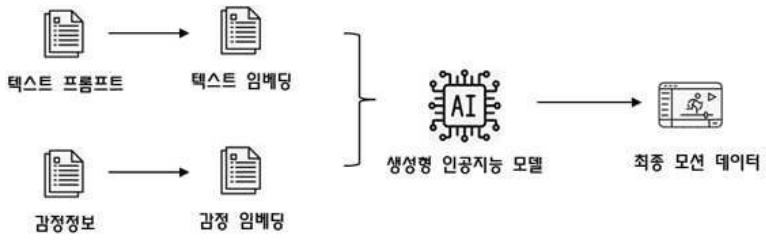
### 도면1



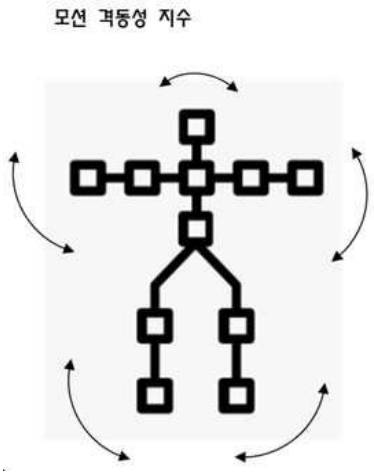

### 도면2



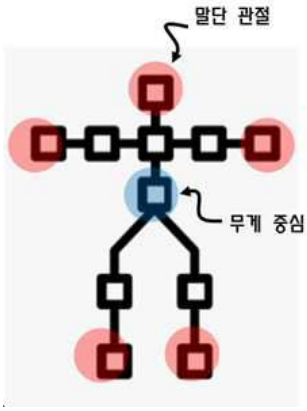

### 도면3



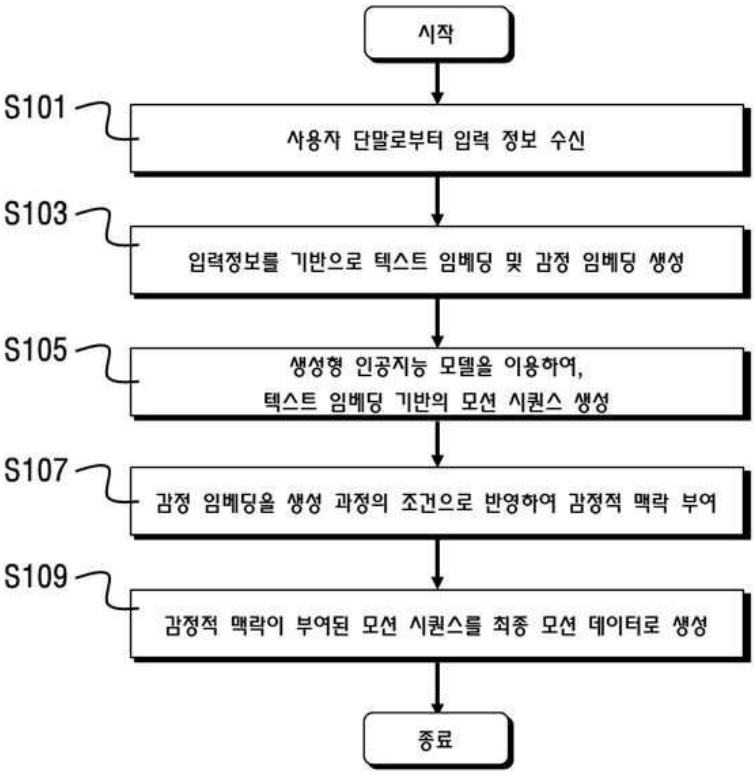
도면4




도면5



도면6


시작
S101
사용자 단말로부터 입력 정보 수신
S103
입력정보를 기반으로 텍스트 임베딩 및 감정 임베딩 생성
S105
생성형 인공지능 모델을 이용하여,
텍스트 임베딩 기반의 모션 시퀀스 생성
S107
감정 임베딩을 생성 과정의 조건으로 반영하여 감정적 맥락 부여
S109
감정적 맥락이 부여된 모션 시퀀스를 최종 모션 데이터로 생성
종료