(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0081116
(43) 공개일자 2025년06월05일

(51) 국제특허분류(Int. Cl.)
G06T 3/10 (2024.01) *G06F 16/9535* (2019.01)
G06T 19/20 (2011.01) *G06V 40/16* (2022.01)
(52) CPC특허분류
G06T 3/10 (2024.01)
G06F 16/9535 (2019.01)
(21) 출원번호 10-2023-0169104
(22) 출원일자 2023년11월29일
심사청구일자 2023년11월29일

(71) 출원인
인하대학교 산학협력단
인천광역시 미추홀구 인하로 100(용현동, 인하대학교)
(72) 발명자
송병철
서울특별시 양천구 목동서로 38, 125동 307호 (목동, 목동신시가지아파트1단지)
김성호
인천광역시 미추홀구 인주대로224번길 6, 406호 (용현동)
(74) 대리인
양성보

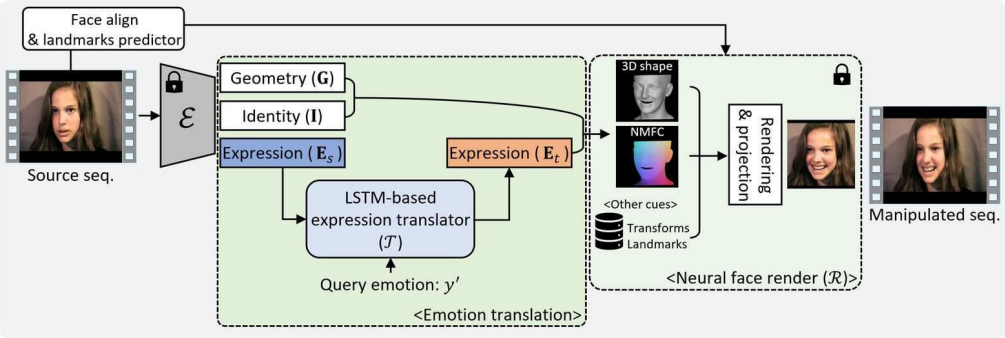
전체 청구항 수 : 총 7 항

(54) 발명의 명칭 **얼굴 표정 조작 모델의 학습을 위한 자기 주의 기반 분별 학습 및 감정 세기에 따른 대조 학습**

(57) 요 약

얼굴 표정 조작 모델의 학습을 위한 자기 주의 기반 분별 학습 및 감정 세기에 따른 대조 학습 기술이 개시된다. 일 실시예에 따른 표정 조작 시스템에 의해 수행되는 얼굴 표정 조작 방법은, 얼굴 표정 조작 모델을 통해 얼굴 영상 정보로부터 신원 정보와 표정 정보를 분리하는 단계; 상기 분리된 표정 정보를 이용하여 목표 감정에 대응하는 표정 벡터를 생성하는 단계; 및 상기 분리된 신원 정보에 대한 벡터와 상기 생성된 목표 감정에 대응하는 표정 벡터를 이용하여 얼굴 영상 정보로부터 얼굴 표정이 조작된 영상 정보를 생성하는 단계를 포함할 수 있다.

대 표 도 - 도1



(52) CPC특허분류
G06T 19/20 (2013.01)
G06V 40/174 (2022.01)
G06T 2207/20081 (2013.01)
G06T 2207/20084 (2013.01)
G06T 2207/30201 (2013.01)

이 발명을 지원한 국가연구개발사업
과제고유번호 1711179347
과제번호 00155915
부처명 과학기술정보통신부
과제관리(전문)기관명 정보통신기획평가원
연구사업명 정보통신인재양성(교육운영수익)
연구과제명 [Ezbaro][정부+민간] 인공지능융합혁신인재양성(2차년도)
기 여 율 1/2
과제수행기관명 인하대학교 산학협력단
연구기간 2023.01.01 ~ 2023.12.31
이 발명을 지원한 국가연구개발사업
과제고유번호 1711186835
과제번호 2022R1A4A1033549
부처명 과학기술정보통신부
과제관리(전문)기관명 한국연구재단
연구사업명 집단연구지원
연구과제명 [Ezbaro] 인공지능 융합형 메타휴먼 플랫폼 연구실
기 여 율 1/2
과제수행기관명 인하대학교
연구기간 2023.03.01 ~ 2024.02.29

명 세 서

청구범위

청구항 1

표정 조작 시스템에 의해 수행되는 얼굴 표정 조작 방법에 있어서,

얼굴 표정 조작 모델을 통해 얼굴 영상 정보로부터 신원 정보와 표정 정보를 분리하는 단계;

상기 분리된 표정 정보를 이용하여 목표 감정에 대응하는 표정 벡터를 생성하는 단계; 및

상기 분리된 신원 정보에 대한 벡터와 상기 생성된 목표 감정에 대응하는 표정 벡터를 이용하여 얼굴 영상 정보로부터 얼굴 표정이 조작된 영상 정보를 생성하는 단계

를 포함하는 얼굴 표정 조작 방법.

청구항 2

제1항에 있어서,

상기 얼굴 표정 조작 모델은,

분별기를 중심으로 분별기의 앞 단에 감정 세기 별 상호 정보(mutual information) 기반 대조 학습을 수행하고, 분별기를 중심으로 분별기의 뒷 단에 감정 발현 정도에 따라 주의 가중치(attention weight)가 결정되도록 주의 기반 분별 학습을 수행하는 것을 특징으로 하는 얼굴 표정 조작 방법.

청구항 3

제2항에 있어서,

상기 얼굴 표정 조작 모델은,

LSTM 기반의 표정 변환기를 통해 상기 분리된 표정 정보로부터 생성된 목표 감정에 대응하는 표정 벡터에 대해 동일한 감정일지라도 감정 세기에 따라 달라질 수 있는 표정 정보를 학습하는 것을 특징으로 하는 얼굴 표정 조작 방법.

청구항 4

제2항에 있어서,

상기 얼굴 표정 조작 모델은,

자기주의를 통해 감정 학습에 도움이 되는 프레임 정보에 더 집중할 수 있도록 입력 정보를 재정의하는 것을 특징으로 하는 얼굴 표정 조작 방법.

청구항 5

제1항에 있어서,

상기 표정 벡터를 생성하는 단계는,

사용자의 쿼리 감정에 기초하여 목표 감정을 입력받고, LSTM 기반의 표정 변환기를 통해 상기 분리된 표정 정보로부터 상기 입력받은 목표 감정에 대응하는 표정 벡터를 생성하는 단계

를 포함하는 얼굴 표정 조작 방법.

청구항 6

제1항에 있어서,

상기 얼굴 표정이 조작된 영상 정보를 생성하는 단계는,

상기 분리된 신원 정보에 대한 벡터, 상기 생성된 목표 감정에 대응하는 표정 벡터 및 얼굴 표정과 관련된 부가 정보를 이용한 렌더링 및 프로젝션을 통해 상기 얼굴 영상 정보로부터 얼굴 표정이 조작된 영상 정보를 생성하는 단계

를 포함하는 얼굴 표정 조작 방법.

청구항 7

표정 조작 시스템에 있어서,

얼굴 표정 조작 모델을 통해 얼굴 영상 정보로부터 신원 정보와 표정 정보를 분리하는 영상 분리부;

상기 분리된 표정 정보를 이용하여 목표 감정에 대응하는 표정 벡터를 생성하는 표정 벡터 생성부; 및

상기 분리된 신원 정보에 대한 벡터와 상기 생성된 목표 감정에 대응하는 표정 벡터를 이용하여 얼굴 영상 정보로부터 얼굴 표정이 조작된 영상 정보를 생성하는 영상 생성부

를 포함하는 표정 조작 시스템.

발명의 설명

기 술 분 야

[0001] 아래의 설명은 얼굴 표정 조작 기술에 관한 것이다.

배 경 기 술

[0003] 가장 최근, NED는 LSTM을 통해 쿼리 감정(query emotion) 정보를 시퀀스 프레임(sequence frame)들에 골고루 반영하는 비디오-투-비디오(video-to-video) 얼굴 표정 조작(Facial Expression Manipulation; FEM) 프레임워크를 제안하였다. 이 연구는 얼굴 표정 조작 태스크에서 비디오-투-비디오 감정 변환(video-to-video emotion translation)을 최초로 정립시켰다는 점에서 의미가 있다. 이들은 서브젝트(subject)의 신원(identity) 및 표정 벡터(expression vector)를 성공적으로 분리(disentangle)시킨 3차원 지오메트리 기법을 도입하여 서브젝트의 신원이 변하는 문제까지 커버한다.

[0004] 하지만 이러한 관점으로 제안된 기법들은 얼굴 표정 조작 프레임워크 구조의 고도화에만 초점을 맞추고 있으며, 감정 공간에서의 심층적인 분석이 이루어지지 않았다. 그래서, 앞선 기법들에 의해 조작된 이미지(비디오)들은 불명확하게 표정을 표현하기도 한다.

발명의 내용

해결하려는 과제

[0006] 고차원 의미론적 속성인 얼굴 표정을 합성하는 얼굴 표정 조작은 입력 감정에 대응하는 표정을 합성하는 것을 목적으로 한다. 이를 위해, 기존 발표된 연구들은 다양한 감정 정보를 활용하거나 모델의 구조를 고도화 하는 방향으로 성능 개선을 추구하였다. 하지만, 감정이라는 의미론적 속성에 대한 깊이 있는 고찰이 부족하였던 기존 연구들은 결과적으로 만족스럽지 못한 표정 조작으로 이어졌다.

[0007] 본 발명은 이러한 문제를 해결하기 위해 감정 발현 정도에 따라 주의 가중치(attention weight)가 결정되는 주의 기반 분별 학습(attention-based discriminative learning)과 함께 감정 세기 별 상호 정보(mutual information) 기반 대조 학습으로 보다 개선된 감정 학습을 제안한다. 이에, 픽셀 공간(pixel space)에서의 정성적 비교는 물론 잠재 공간(latent space)에서의 임베딩(embedding) 경향을 분석함으로써 기존 기법들과 차별화 된 결과를 보여준다.

과제의 해결 수단

[0009] 표정 조작 시스템에 의해 수행되는 얼굴 표정 조작 방법은, 얼굴 표정 조작 모델을 통해 얼굴 영상 정보로부터 신원 정보와 표정 정보를 분리하는 단계; 상기 분리된 표정 정보를 이용하여 목표 감정에 대응하는 표정 벡터를 생성하는 단계; 및 상기 분리된 신원 정보에 대한 벡터와 상기 생성된 목표 감정에 대응하는 표정 벡터를 이용하여 얼굴 영상 정보로부터 얼굴 표정이 조작된 영상 정보를 생성하는 단계를 포함할 수 있다.

[0010] 상기 얼굴 표정 조작 모델은, 분별기를 중심으로 분별기의 앞 단에 감정 세기 별 상호 정보(mutual information) 기반 대조 학습을 수행하고, 분별기를 중심으로 분별기의 뒷 단에 감정 발현 정도에 따라 주의 가중치(attention weight)가 결정되도록 주의 기반 분별 학습을 수행하는 것일 수 있다.

[0011] 상기 얼굴 표정 조작 모델은, LSTM 기반의 표정 변환기를 통해 상기 분리된 표정 정보로부터 생성된 목표 감정에 대응하는 표정 벡터에 대해 동일한 감정일지라도 감정 세기에 따라 달라질 수 있는 표정 정보를 학습할 수 있다.

[0012] 상기 얼굴 표정 조작 모델은, 자기 주의를 통해 감정 학습에 도움이 되는 프레임 정보에 더 집중할 수 있도록 입력 정보를 재정의할 수 있다.

[0013] 상기 표정 벡터를 생성하는 단계는, 사용자의 쿼리 감정에 기초하여 목표 감정을 입력받고, LSTM 기반의 표정 변환기를 통해 상기 분리된 표정 정보로부터 상기 입력받은 목표 감정에 대응하는 표정 벡터를 생성하는 단계를 포함할 수 있다.

[0014] 상기 얼굴 표정이 조작된 영상 정보를 생성하는 단계는, 상기 분리된 신원 정보에 대한 벡터, 상기 생성된 목표 감정에 대응하는 표정 벡터 및 얼굴 표정과 관련된 부가 정보를 이용한 렌더링 및 프로젝션을 통해 상기 얼굴 영상 정보로부터 얼굴 표정이 조작된 영상 정보를 생성하는 단계를 포함할 수 있다.

[0015] 표정 조작 시스템은, 얼굴 표정 조작 모델을 통해 얼굴 영상 정보로부터 신원 정보와 표정 정보를 분리하는 영상 분리부; 상기 분리된 표정 정보를 이용하여 목표 감정에 대응하는 표정 벡터를 생성하는 표정 벡터 생성부; 및 상기 분리된 신원 정보에 대한 벡터와 상기 생성된 목표 감정에 대응하는 표정 벡터를 이용하여 얼굴 영상 정보로부터 얼굴 표정이 조작된 영상 정보를 생성하는 영상 생성부를 포함할 수 있다.

발명의 효과

[0017] 감정 별로 구분된 영상을 생성함과 동시에 각 감정을 확실히 표현할 수 있는 영상을 생성할 수 있다.

도면의 간단한 설명

[0019] 도 1은 일 실시예에 있어서, 얼굴 표정 조작 모델을 이용한 얼굴 표정 조작 동작을 설명하기 위한 도면이다.

도 2는 일 실시예에 있어서, 얼굴 표정 조작 모델의 학습 동작을 설명하기 위한 도면이다.

도 3은 일 실시예에 있어서, 표정 조작 시스템의 구성을 설명하기 위한 블록도이다.

도 4는 일 실시예에 있어서, 얼굴 표정 조작 방법을 설명하기 위한 흐름도이다.

도 5는 일 실시예에 있어서, 얼굴 표정 조작 모델의 기술적 효과를 설명하기 위한 예이다.

발명을 실시하기 위한 구체적인 내용

[0020] 이하, 실시예를 첨부한 도면을 참조하여 상세히 설명한다.

[0022] 메타버스 및 소셜 로봇 분야에서 얼굴 표정 조작 모델은 중요한 연구로 여겨진다. 하지만, 감정이라는 의미론적 속성의 표현(representation)을 고차원 픽셀 공간(pixel space)에서 재구성 해야하며 다양한 주관적 관점에서 해석될 수 있기 때문에 도적적이다. 이러한 문제 상황에도 불구하고 기존의 연구들은 단순히 기술의 고도화에만 집중한다. 이에 따라, 얼굴 표정 조작 연구는 영상 간 변환을 넘어 동영상 간 변환까지 빠른 발전 속도를

보여주고 있으나 감정의 표현 측면에서의 해석은 아직 부족하다. 결국, 얼굴 영상에서 원하는 표정을 조작 한 결과는 아직 미흡한 결과를 보여주고 있다.

[0023] 실시예에서는 감정 표현 학습의 효과를 극대화하기 위해 감정 발현 정도에 따라 주의 가중치가 결정되는 주의 기반 분별 학습과 감정 세기 별 상호 정보 기반 대조 학습으로 보다 개선된 감정 학습을 수행하는 동작에 대하여 설명하기로 한다.

[0024] 도 1은 일 실시예에 있어서, 얼굴 표정 조작 모델을 이용한 얼굴 표정 조작 동작을 설명하기 위한 도면이다.

[0025] 표정 조작 시스템은 Face align & landmarks predictor의 경우 FAN(Facial Alignment Network)을 이용하여 얼굴의 복수 개(예를 들면, 68개)의 랜드마크(landmark)를 추출하고 정면 얼굴 템플릿(template) 정보와 비교하여 정면 얼굴로 정렬할 수 있는 얼라인먼트 계수(alignment coefficient)를 반환할 수 있다. 이때, 도 1에서 표시된 암호화의 경우, 뉴럴 네트워크 기반 모듈이지만 실제 그레디언트(gradient)를 통한 업데이트가 진행되지 않는 것을 의미한다. 표정 조작 시스템은 반환된 얼라인먼크 계수를 뉴럴 얼굴 렌더(Neural face render)를 통한 얼굴 영상 생성에 이용할 수 있다.

[0026] 표정 조작 시스템은 얼굴 표정 조작 모델을 통해 얼굴 영상 정보로부터 신원 정보와 표정 정보를 분리할 수 있다. 또한, 표정 조작 시스템은 얼굴 영상 정보로부터 지오메트리 정보를 추출할 수 있다. 예를 들면, 표정 조작 시스템은 DECA라는 파라메트릭 3D 변형 가능한 모델(parametric 3D morphable model)의 인코더를 통해 얼굴 영상 정보로부터 지오메트리 정보를 추출할 수 있다. 여기서, 지오메트리 정보에는 포즈(pose), 알베도(albedo), 조도(illuminance) 등이 포함될 수 있다. 알베도와 조도 정보들은 영상의 정확한 색의 매핑을 위해 필요한 정보들을 의미한다. 표정 조작 시스템은 분리된 표정 정보에 대한 잠재 벡터만 이용하여 목표 감정에 대응하는 표정 벡터를 생성(조작)할 수 있다. 표정 조작 시스템은 사용자의 쿼리 감정에 기초하여 목표 감정을 입력받고, LSTM 기반의 표정 변환기를 통해 분리된 표정 정보로부터 입력받은 목표 감정에 대응하는 표정 벡터를 생성할 수 있다. 표정 조작 시스템은 분리된 신원 정보에 대한 벡터, 생성된 목표 감정에 대응하는 표정 벡터 및 이외의 정보들을 이용하여 얼굴 영상 정보로부터 얼굴 표정이 조작된 영상 정보를 생성할 수 있다. 예를 들면, 이외의 정보들로 3D shape, NMFC, transforms 및 landmarks이 포함될 수 있다. 3D shape란 3차원 변경 가능한 모델을 이용하여 생성된 3차원 헤드(head)를 의미한다. NMFC는 앞서 언급한 색 정보와 관련있는 것으로, 3D shape중 음영이 표현되어야 하는 정보를 전달한다. transforms와 landmarks는 앞서 설명한 얼라인먼트 정보들을 이용하여 원본 영상의 얼굴 방향을 복원하기 위해 이용될 수 있다. 이와 같은 방법을 이용하는 이유는 3D 변형 가능한 모델이 섬세한 조작에 유리하기 때문이다. 여기서 섬세한 조작이란 입꼬리를 올리는 정도, 미간을 찡그린 정도를 조작할 수 있음을 의미한다. 표정 조작 시스템은 분리된 신원 정보에 대한 벡터, 생성된 목표 감정에 대응하는 표정 벡터 및 얼굴 표정과 관련된 부가 정보를 이용한 렌더링 및 프로젝션을 통해 얼굴 영상 정보로부터 얼굴 표정이 조작된 영상 정보를 생성할 수 있다. 표정 조작 시스템은 3D 정보들이 들어오면 뉴럴 네트워크(neural network) 기반의 모듈을 이용하여 2D 프로젝션(projection) 영상을 생성할 수 있다. 이때, 현재 프레임을 기준으로 이전 프레임(바로 앞 프레임)의 프로젝션 정보가 현재 프레임의 프로젝션에 이용될 수 있다. 이러한 재귀적 구조를 통해 자연스러운 동영상 생성이 가능하다.

[0027] 도 2는 일 실시예에 있어서, 얼굴 표정 조작 모델의 학습 동작을 설명하기 위한 도면이다.

[0028] 표정 조작 시스템은 분별기를 중심으로 분별기의 앞 단에 감정 세기 별 상호 정보(mutual information) 기반 대조 학습을 수행하고, 분별기를 중심으로 분별기의 뒷 단에 감정 발현 정도에 따라 주의 가중치(attention weight)가 결정되도록 주의 기반 분별 학습을 수행할 수 있다.

[0029] 얼굴 표정 조작 모델은 LSTM 기반의 표정 변환기를 통해 분리된 표정 정보로부터 생성된 목표 감정에 대응하는 표정 벡터에 대해 동일한 감정일지라도 감정 세기에 따라 달라질 수 있는 표정 정보를 학습함으로써 강한 감정에서 특징적으로 발현되어야 하는 정보를 학습할 수 있다.

[0030] 또한, 얼굴 표정 조작 모델은 자기주의를 통해 감정 학습에 도움이 되는 프레임 정보에 더 집중할 수 있도록 입력 정보를 재정의할 수 있다.

[0031] 어텐션 모듈의 경우 표정 벡터들을 이용해 쿼리(query), 키(key), 밸류(value)를 정의한 후 쿼리와 키의 유사성(similarity)을 측정한다. 이때, 쿼리는 맥시멈 감정 강도(maximum emotional intensity)를 나타내는 프레임이다. 따라서 가장 감정 정보를 잘 나타내는 정보와의 유사성을 통해 주의를 가해야 하는 프레임 정보를 추출 및 적용할 수 있다. 이렇게 중요도에 따른 가중치가 적용된 표정 벡터 시퀀스가 판별자(분별기) D의 입력으로 사용되어 모델이 라벨에 해당하는 감정 정보가 확실한 프레임들의 정보를 더욱 분명하게 학습할 수 있게 된다.

[0032] 도 3은 일 실시예에 있어서, 표정 조작 시스템의 구성을 설명하기 위한 블록도이고, 도 4는 일 실시예에 있어서, 얼굴 표정 조작 방법을 설명하기 위한 흐름도이다.

[0033] 표정 조작 시스템(100)의 프로세서는 영상 분리부(310), 표정 벡터 생성부(320) 및 영상 생성부(330)를 포함할 수 있다. 이러한 프로세서의 구성요소들은 표정 조작 시스템에 저장된 프로그램 코드가 제공하는 제어 명령에 따라 프로세서에 의해 수행되는 서로 다른 기능들(different functions)의 표현들일 수 있다. 프로세서 및 프로세서의 구성요소들은 도 4의 얼굴 표정 조작 방법이 포함하는 단계들(410 내지 430)을 수행하는 표정 조작 시스템을 제어할 수 있다. 이때, 프로세서 및 프로세서의 구성요소들은 메모리가 포함하는 운영체제의 코드와 적어도 하나의 프로그램의 코드에 따른 명령(instruction)을 실행하도록 구현될 수 있다.

[0034] 프로세서는 얼굴 표정 조작 방법을 위한 프로그램의 파일에 저장된 프로그램 코드를 메모리에 로딩할 수 있다. 예를 들면, 표정 조작 시스템에서 프로그램이 실행되면, 프로세서는 운영체제의 제어에 따라 프로그램의 파일로부터 프로그램 코드를 메모리에 로딩하도록 표정 조작 시스템을 제어할 수 있다. 이때 프로세서는 영상 분리부(310), 표정 벡터 생성부(320) 및 영상 생성부(330) 각각에 메모리에 로딩된 프로그램 코드 중 대응하는 부분의 명령을 실행하여 이후 단계들(410 내지 430)을 실행하기 위한 프로세서의 서로 다른 기능적 표현들일 수 있다.

[0035] 단계(410)에서 영상 분리부(310)는 얼굴 표정 조작 모델을 통해 얼굴 영상 정보로부터 신원 정보와 표정 정보를 분리할 수 있다.

[0036] 단계(420)에서 표정 벡터 생성부(320)는 분리된 표정 정보를 이용하여 목표 감정에 대응하는 표정 벡터를 생성할 수 있다. 표정 벡터 생성부(320)는 사용자의 쿼리 감정에 기초하여 목표 감정을 입력받고, LSTM 기반의 표정 변환기를 통해 분리된 표정 정보로부터 입력받은 목표 감정에 대응하는 표정 벡터를 생성할 수 있다.

[0037] 단계(430)에서 영상 생성부(330)는 분리된 신원 정보에 대한 벡터와 생성된 목표 감정에 대응하는 표정 벡터를 이용하여 얼굴 영상 정보로부터 얼굴 표정이 조작된 영상 정보를 생성할 수 있다. 영상 생성부(330)는 분리된 신원 정보에 대한 벡터, 생성된 목표 감정에 대응하는 표정 벡터 및 얼굴 표정과 관련된 부가 정보를 이용한 렌더링 및 프로젝션을 통해 얼굴 영상 정보로부터 얼굴 표정이 조작된 영상 정보를 생성할 수 있다. 이에, 감정 표현의 정확도를 향상시킨 영상을 생성할 수 있다.

[0038] 도 5는 일 실시예에 있어서, 얼굴 표정 조작 모델의 기술적 효과를 설명하기 위한 예이다.

[0039] 표 1은 영상 생성의 정량적 성능을 비교한 표이다.

[0040]

	ICface [55]		GANmut [28]		NED [4]		MECA	
	FID	FED	FID	FED	FID	FED	FID	FED
Happy	58.2	40.9	21.6	**2.1**	15.3	5.0	**12.1**	3.5
Sad	65.7	59.9	23.1	3.4	12.9	4.8	**10.3**	**2.8**
Surprised	58.3	57.4	31.5	4.5	12.5	5.5	**11.4**	**3.0**
Fear	63.3	59.4	30.6	4.0	11.1	**3.1**	**9.8**	**3.1**
Disgusted	70.3	63.5	29.1	3.2	12.0	3.2	**10.1**	**3.0**
Angry	67.5	58.7	33.6	6.3	13.1	3.4	**11.1**	**2.7**
Avg.	63.9	56.6	28.3	3.9	12.8	4.2	**10.8**	**3.0**

[0041] FID와 FED는 각각 신원 보존과 감정 표현 능력에 대한 지표이다. 두가지 지표는 낮은 값일 수록 좋은 성능이라고 간주될 수 있다. FID 측면에서 실시예에서 제안된 방법(MECA)은 모든 감정 측면에서 우수한 결과를 보여준다. 얼굴 영상을 조작하는 연구에서 신원의 보존은 매우 중요한 요소로써 해석될 수 있다. 이러한 관점에서 발명 모델의 성능은 매우 우수하다고 판단될 수 있다. 또한, FED의 비교는 감정 표현의 직관적인 지표로 작용될 수 있기 때문에 매우 중요하다. 실시예에서 제안된 방법(MECA)은 비교 기법 대비 최대 60.5 좋은 성능을 보여준다. 이는 감정이라는 의미론적 속성의 학습을 위해 고안된 발명 기법이 충분한 효과를 보였다고 해설될 수 있다.

[0042] 도 5를 참고하면, 감정 별로 표현해야 되는 특징이 잘 표현된 것을 볼 수 있다. 실시예에서 제안된 방법(MECA)은 감정 별로 구분된 영상을 생성함과 동시에 각 감정을 확실히 표현할 수 있는 영상이 생성된 것을 확인할 수 있다. 이러한 결과는 이번 발명이 추구하는 방향과 일치한다.

[0043] 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), 마이크로컴

퓨터, 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.

[0044] 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.

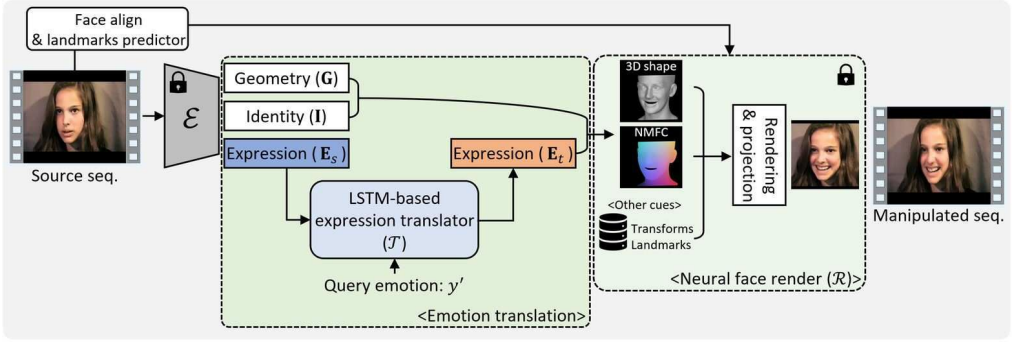
[0045] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic 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.

[0046] 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.

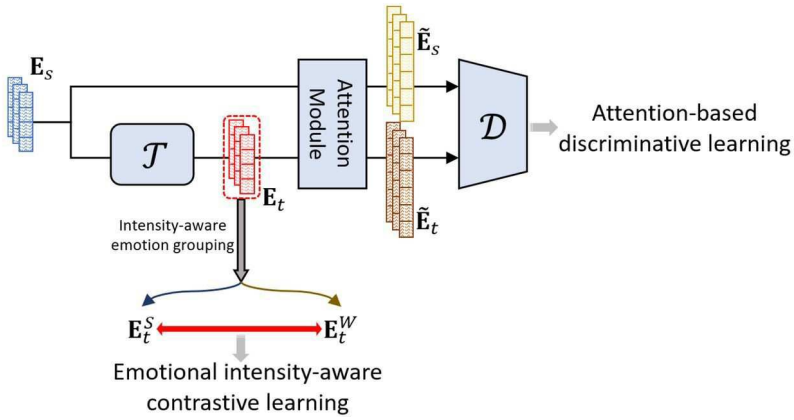
[0047] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

도면

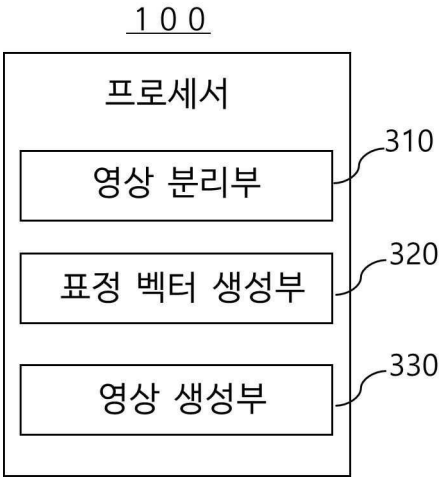
도면1



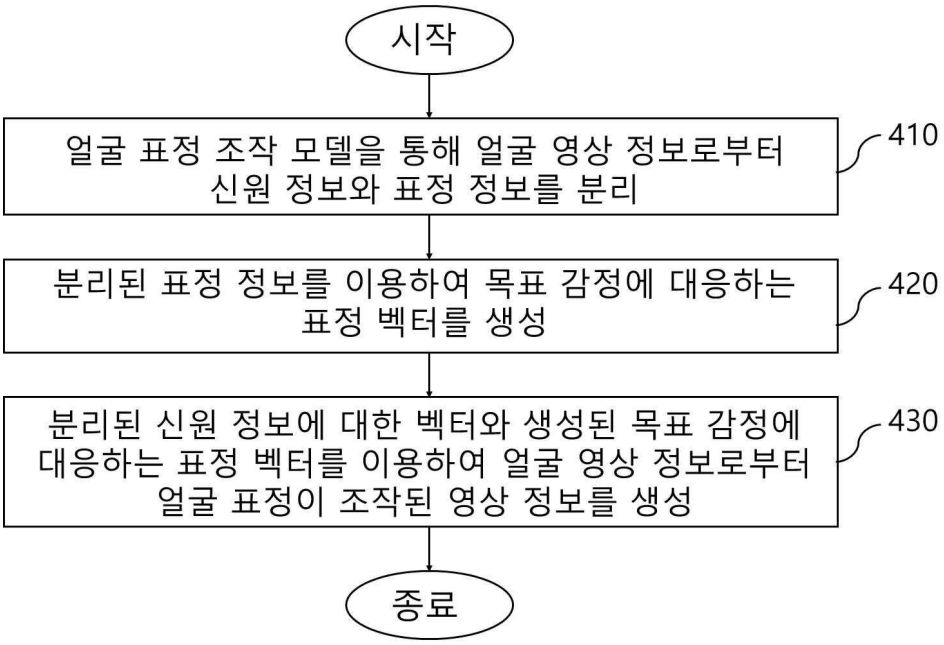
도면2


$\mathbf{E}_s$
$\mathcal{T}$
$\mathbf{E}_t$
Attention Module
$\tilde{\mathbf{E}}_s$
$\tilde{\mathbf{E}}_t$
$\mathcal{D}$
Attention-based discriminative learning
Intensity-aware emotion grouping
$\mathbf{E}_t^S$
$\mathbf{E}_t^W$
Emotional intensity-aware contrastive learning

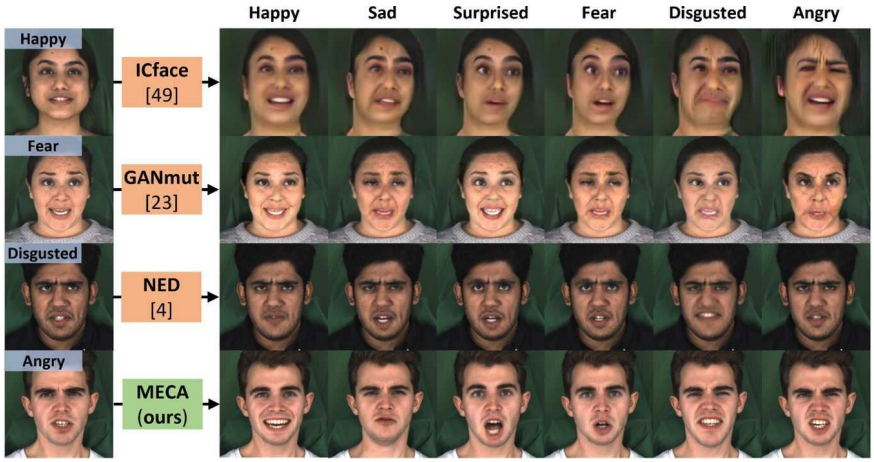
도면3


100
프로세서
영상 분리부
310
표정 벡터 생성부
320
영상 생성부
330

도면4


시작
얼굴 표정 조작 모델을 통해 얼굴 영상 정보로부터 신원 정보와 표정 정보를 분리
410
분리된 표정 정보를 이용하여 목표 감정에 대응하는 표정 벡터를 생성
420
분리된 신원 정보에 대한 벡터와 생성된 목표 감정에 대응하는 표정 벡터를 이용하여 얼굴 영상 정보로부터 얼굴 표정이 조작된 영상 정보를 생성
430
종료

도면5


Happy
Sad
Surprised
Fear
Disgusted
Angry
Happy
Fear
Disgusted
Angry
ICface [49]
GANmut [23]
NED [4]
MECA (ours)