



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0047779
(43) 공개일자 2024년04월12일

- | | |
|---|---|
| <p>(51) 국제특허분류(Int. Cl.)
 <i>G06T 13/40</i> (2011.01) <i>G06T 17/20</i> (2006.01)
 <i>G06T 7/11</i> (2017.01) <i>G06T 7/70</i> (2017.01)
 <i>G06V 40/16</i> (2022.01) <i>G10L 15/02</i> (2006.01)</p> <p>(52) CPC특허분류
 <i>G06T 13/40</i> (2013.01)
 <i>G06T 17/20</i> (2013.01)</p> <p>(21) 출원번호 10-2022-0127246
 (22) 출원일자 2022년10월05일
 심사청구일자 2022년10월05일</p> | <p>(71) 출원인
 울산과학기술원
 울산광역시 울주군 언양읍 유니스트길 50</p> <p>(72) 발명자
 백승렬
 울산광역시 울주군 언양읍 유니스트길 50
 김은서
 울산광역시 울주군 언양읍 유니스트길 50</p> <p>(74) 대리인
 전용준</p> |
|---|---|

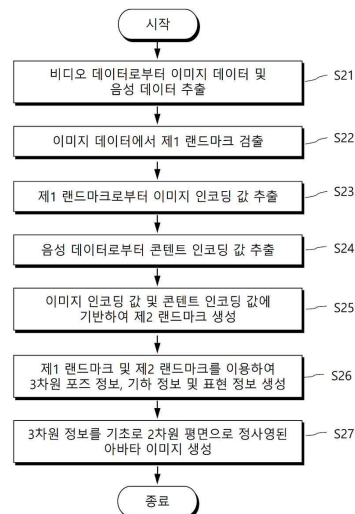
전체 청구항 수 : 총 15 항

(54) 발명의 명칭 **아바타 생성 방법 및 그 시스템**

(57) 요약

아바타를 생성하는 방법 및 그 시스템을 개시한다. 본 방법은 비디오 데이터로부터 이미지 데이터 및 음성 데이터를 추출하고, 이미지 데이터로부터 검출된 제1 랜드마크(landmark)로부터 이미지 인코딩 값을 추출하며, 음성 데이터로부터 콘텐츠 인코딩 값을 추출하고, 이미지 인코딩 값과 콘텐츠 인코딩 값에 기반하여 제2 랜드마크를 추출하고, 제 1 랜드마크 및 제 2 랜드마크를 이용하여 3차원의 포즈 정보, 기하 정보 및 표현 정보를 생성하고, 3차원의 포즈 정보, 기하 정보 및 표현 정보를 기초로 2차원 평면으로 정사영된 아바타 이미지를 생성한다.

대표도 - 도2



(52) CPC특허분류

G06T 7/11 (2017.01)

G06T 7/70 (2017.01)

G10L 15/02 (2013.01)

G06T 2207/20081 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711159656
과제번호	2021-0-01778-002
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	차세대인공지능핵심원천기술개발
연구과제명	실제와 지각 역치 이하 수준까지 동일한 특성을 갖는 인물 영상 합성 기술 및 판별
기술 개발	
기 여 율	1/1
과제수행기관명	울산과학기술원
연구기간	2022.02.01 ~ 2023.01.31

명세서

청구범위

청구항 1

전자 장치가 아바타(avatar)를 생성하는 방법에 있어서,
비디오 데이터로부터 이미지 데이터 및 음성 데이터를 추출하는 단계;
상기 이미지 데이터로부터 검출된 제1 랜드마크(landmark)로부터 이미지 인코딩 값을 추출하는 단계;
상기 음성 데이터로부터 콘텐츠 인코딩 값을 추출하는 단계;
상기 이미지 인코딩 값과 상기 콘텐츠 인코딩 값에 기반하여 제2 랜드마크를 추출하는 단계;
상기 제1 랜드마크 및 상기 제2 랜드마크를 이용하여 3차원의 포즈 정보, 기하 정보 및 표현 정보를 생성하는 단계; 및
상기 3차원의 포즈 정보, 기하 정보 및 표현 정보를 기초로 2차원 평면으로 정사영된 아바타 이미지를 생성하는 단계를 포함하는 방법.

청구항 2

제 1 항에 있어서,
상기 3차원의 포즈 정보, 기하 정보 및 표현 정보를 생성하는 단계는,
상기 제1 랜드마크를 기 학습된 3DMM(3D Morphable face Model) 기반의 신경망에 입력하여 포즈 파라미터로 변환하는 단계; 및
상기 제2 랜드마크를 상기 기 학습된 3DMM 기반의 신경망에 입력하여 기하 파라미터 및 표현 파라미터로 변환하는 단계를 포함하는 것인, 방법.

청구항 3

제 1 항에 있어서,
상기 아바타 이미지를 생성하는 단계는,
상기 3차원의 포즈 정보, 기하 정보 및 표현 정보로부터 3차원 메쉬(mesh)를 생성하는 단계;
상기 3차원 메쉬의 특징점을 2차원 평면으로 정사영하여 제3 랜드마크를 추출하는 단계; 및
상기 제3 랜드마크를 기 학습된 cGAN(Conditional Generative Adversarial Network) 기반의 신경망에 입력하여 상기 아바타 이미지를 생성하는 단계를 포함하는 것인 방법.

청구항 4

제 3 항에 있어서,
상기 제3 랜드마크를 추출하는 단계는,
하기 수식 1에 따라 상기 제3 랜드마크를 추출하는 것인 방법:
[수식 1]

$$l_p = s_t * Pr * R_t * L_p^{3d} + t_t,$$

여기서 (R_t , t_t , S_t)는 포즈 파라미터의 값이며, Pr 은 정사영 행렬 ($Pr = \{(1,0,0), (0,1,0)\}$), L_p^{3d} 는 3차원

메쉬의 특징점이다.

청구항 5

제 1 항에 있어서,

상기 제1 랜드마크로부터 이미지 인코딩 값을 추출하는 단계는,

상기 이미지 데이터로부터 얼굴 영역을 추출하는 단계; 및

상기 얼굴 영역을 기 학습된 신경망에 입력하여 상기 제1 랜드마크를 검출하는 단계를 더 포함하는 것인, 방법.

청구항 6

제 5 항에 있어서,

상기 제1 랜드마크로부터 이미지 인코딩 값을 추출하는 단계는,

상기 제1 랜드마크를 기 학습된 MLP(Multi-Layer Perceptron) 기반의 신경망에 입력하여 상기 이미지 인코딩 값을 추출하는 단계를 더 포함하는 것인, 방법.

청구항 7

제 1 항에 있어서,

상기 제2 랜드마크를 추출하는 단계는,

상기 이미지 인코딩 값에 상기 콘텐츠 인코딩 값을 감산, 가산, AND, OR, XOR, NOT 및 쉬프팅(Shifting) 중 적어도 하나의 논리 연산을 수행하여 상기 제2 랜드마크를 추출하는 것인, 방법.

청구항 8

제 1 항에 있어서,

상기 제2 랜드마크를 추출하는 단계는,

상기 이미지 인코딩 값과 상기 콘텐츠 인코딩 값을 기 학습된 MLP 기반의 신경망에 입력하여 제2 랜드마크로 변환하는 것인, 방법.

청구항 9

적어도 하나의 클라이언트 장치와 통신하는 서버가 아바타를 생성하는 방법에 있어서,

상기 적어도 하나의 클라이언트 장치로부터 비디오 데이터를 수신하는 단계;

상기 비디오 데이터로부터 이미지 데이터 및 음성 데이터를 추출하는 단계;

상기 이미지 데이터로부터 검출된 제1 랜드마크(landmark)로부터 이미지 인코딩 값을 추출하는 단계;

상기 음성 데이터로부터 콘텐츠 인코딩 값을 추출하는 단계;

상기 이미지 인코딩 값과 상기 콘텐츠 인코딩 값에 기반하여 제2 랜드마크를 추출하는 단계;

상기 제1 랜드마크 및 상기 제2 랜드마크를 이용하여 3차원의 포즈 정보, 기하 정보 및 표현 정보를 생성하는 단계;

상기 3차원의 포즈 정보, 기하 정보 및 표현 정보를 기초로 2차원 평면으로 정사영된 아바타 이미지를 생성하는 단계; 및

상기 생성된 아바타 이미지를 상기 클라이언트 장치로 전송하는 단계를 포함하는 방법.

청구항 10

제 9 항에 있어서,

상기 3차원의 포즈 정보, 기하 정보 및 표현 정보를 생성하는 단계는,

상기 제1 랜드마크를 기 학습된 3DMM(3D Morphable face Model) 기반의 신경망에 입력하여 포즈 파라미터로 변환하는 단계; 및

상기 제2 랜드마크를 상기 기 학습된 3DMM 기반의 신경망에 입력하여 기하 파라미터 및 표현 파라미터로 변환하는 단계를 포함하는 것인, 방법.

청구항 11

제 9 항에 있어서,

상기 아바타 이미지를 생성하는 단계는,

상기 3차원의 포즈 정보, 기하 정보 및 표현 정보로부터 3차원 메쉬(mesh)를 생성하는 단계;

상기 3차원 메쉬의 특징점을 2차원 평면으로 정사영하여 제3 랜드마크를 추출하는 단계; 및

상기 제3 랜드마크를 기 학습된 cGAN(Conditional Generative Adversarial Network) 기반의 신경망에 입력하여 상기 아바타 이미지를 생성하는 단계를 포함하는 것인 방법.

청구항 12

제 9 항에 있어서,

상기 제1 랜드마크로부터 이미지 인코딩 값을 추출하는 단계는,

상기 이미지 데이터로부터 얼굴 영역을 추출하는 단계; 및

상기 얼굴 영역을 기 학습된 신경망에 입력하여 상기 제1 랜드마크를 검출하는 단계를 더 포함하는 것인, 방법.

청구항 13

제 9 항에 있어서,

상기 클라이언트 장치로부터 비디오 데이터를 수신하는 단계는,

상기 클라이언트 장치에 의해 비디오 데이터에서 얼굴 영역이 검출된 얼굴 비디오 데이터를 수신하는 것인 방법.

청구항 14

하나 이상의 인스트럭션을 저장하는 메모리; 및

상기 메모리에 저장된 상기 하나 이상의 인스트럭션을 실행하는 프로세서를 포함하고,

상기 프로세서는 상기 하나 이상의 인스트럭션을 실행함으로써,

제1 항의 방법을 수행하는, 전자 장치.

청구항 15

하드웨어인 컴퓨터와 결합되어, 제1 항의 방법을 수행할 수 있도록 컴퓨터에서 독출가능한 기록매체에 저장된 컴퓨터프로그램.

발명의 설명

기술 분야

[0001] 본 발명은 아바타 생성 방법 및 그 장치에 관한 것으로, 보다 상세하게는, 발화에 따라 얼굴 형상이 자연스럽게 변경되는 아바타를 생성하는 방법 및 그 장치에 관한 것이다.

배경 기술

[0002] 가상현실은 사용자가 HMD(Head Mounted Display)장치 등의 입출력장치를 착용함으로써 컴퓨터 상에서 구현되는 가상의 공간을 체험할 수 있게 하는 기술로서, 사용자가 현실 세계와 유사하게 구현되는 가상의 공간에서 시간

적, 공간적 제약없이 특정 상황에 대한 체험을 할 수 있도록 하는 기술이다. 이러한 가상현실 기술은 값비싼 장비 및 가상현실 콘텐츠의 부족으로 인해 빠른 시일 내에 대중화되기 어려울 것으로 전망되기도 했으나, 최근에는 VR 및 AR장치의 보급과 비대면 콘텐츠에 대한 수요가 증가함에 따라, 가상현실 기술 또한 재조명 받고 있는 상황이다.

[0003] 메타버스 기술이란 가상 자아인 아바타를 통해 경제, 사회, 문화, 정치 활동 등을 이어가는 4차원 가상 시공간 할 수 있다. 이러한 메타버스 기술은 가상현실, 증강현실, 거울세계 및 라이프로그를 포함하는 개념이다. 이러한 메타버스 기술은 게임, 교육, 쇼핑, 의료 및 관광 등 다양한 기술분야와 접목되어 활용되고 있고, 구체적으로 메타버스에서 구현한 게임, 교육, 쇼핑, 의료 및 관광 등의 플랫폼을 아바타를 통해 체험할 수 있어 아바타 생성은 필수적이며, 이러한 플랫폼들이 활성화됨에 따라 메타버스 내에서 타인과 소통할 수 있는 기회가 증가하고 있다. 그러나, 종래의 아바타 생성 기술은 단순히 사용자의 외형과 유사한 아바타를 생성하고 이를 통하여 사용자의 메시지를 전달하는데 그치고 있으며, 아바타를 통하여 그 메시지에 따른 감정 상태를 전달함에는 어려움이 있다.

[0004] 따라서, 사용자 메시지를 얼굴 형상의 움직임과 연결하여 보다 효율적으로 아바타를 통해 사용자 상태를 표현함으로써 다른 사용자와의 상호작용을 향상시킬 수 있도록 하는 아바타를 생성하는 방법 및 그 시스템의 개발이 필요한 실정이다.

선행기술문헌

특허문헌

[0005] (특허문헌 0001) 미국 특허 공개공보 제 2021-0343061 호 ("ANIMATED EXPRESSIVE ICON")

발명의 내용

해결하려는 과제

[0006] 본 발명은 상술한 문제점을 해결하기 위한 것으로, 본 발명의 목적은 메타버스 플랫폼이나 화상회의 시스템에서 활용 가능한 아바타 생성하는데 있으며, 보다 구체적으로 이미지 데이터와 음성 데이터를 연결하여 발화하는 비디오 데이터를 생성하는데 있다.

[0007] 다만, 본 실시예가 이루고자 하는 기술적 과제는 상기된 바와 같은 기술적 과제들로 한정되지 않으며, 또 다른 기술적 과제들이 더 존재할 수 있다.

과제의 해결 수단

[0008] 상술한 목적을 달성하기 위한 본 발명의 일 실시 예에 따른 아바타를 생성하는 전자 장치의 동작 방법은 비디오 데이터로부터 이미지 데이터 및 음성 데이터를 추출하는 단계; 상기 이미지 데이터로부터 검출된 제1 랜드마크 (landmark)로부터 이미지 인코딩 값을 추출하는 단계; 상기 음성 데이터로부터 콘텐츠 인코딩 값을 추출하는 단계; 상기 이미지 인코딩 값과 상기 콘텐츠 인코딩 값에 기반하여 제2 랜드마크를 추출하는 단계; 상기 제1 랜드마크 및 상기 제2 랜드마크를 이용하여 3차원의 포즈 정보, 기하 정보 및 표현 정보를 생성하는 단계; 및 상기 3차원의 포즈 정보, 기하 정보 및 표현 정보를 기초로 2차원 평면으로 정사영된 아바타 이미지를 생성하는 단계를 개시한다.

[0009] 본 발명의 기타 구체적인 사항들은 상세한 설명 및 도면들에 포함되어 있다.

발명의 효과

[0010] 이상과 같은 다양한 실시 예들에 따르면, 아바타 생성 시스템은 사용자의 발화에 따른 얼굴 형상 변화를 이용하여 아바타를 생성함으로써 사용자의 메시지뿐 아니라 사용자의 감정 상태를 보다 효율적으로 전달할 수 있다.

[0011] 본 발명의 효과들은 이상에서 언급된 효과로 제한되지 않으며, 언급되지 않은 또 다른 효과들은 아래의 기재로부터 통상의 기술자에게 명확하게 이해될 수 있을 것이다.

도면의 간단한 설명

- [0012] 도 1은 본 발명의 일 실시예에 따른 아바타 생성 시스템을 도시한 도면이다.
- 도 2는 본 발명의 일 실시예에 따른 아바타 생성 시스템의 동작 방법을 도시한 순서도이다.
- 도 3은 본 발명의 일 실시예에 따른 MLP를 도시한 일 예이다.
- 도 4는 3차원 정보로부터 아바타 이미지를 생성하는 방법을 구체적으로 도시한 순서도이다.
- 도 5는 본 발명의 일 실시예에 따라 도 2 내지 도 4의 동작을 서버가 아닌 전자 장치가 수행하는 아바타 생성 시스템을 도시한 일 예이다.
- 도 6은 본 발명의 일 실시예에 따른 아바타 생성 방법을 도시한 도면이다.
- 도 7은 본 발명의 일 실시 예에 따른 서버의 구성도이다.

발명을 실시하기 위한 구체적인 내용

- [0013] 본 발명의 이점 및 특징, 그리고 그것들을 달성하는 방법은 첨부되는 도면과 함께 상세하게 후술되어 있는 실시예들을 참조하면 명확해질 것이다. 그러나, 본 발명은 이하에서 개시되는 실시예들에 제한되는 것이 아니라 서로 다른 다양한 형태로 구현될 수 있으며, 단지 본 실시예들은 본 발명의 개시가 완전하도록 하고, 본 발명이 속하는 기술 분야의 통상의 기술자에게 본 발명의 범주를 완전하게 알려주기 위해 제공되는 것이며, 본 발명은 청구항의 범주에 의해 정의될 뿐이다.
- [0014] 본 명세서에서 사용된 용어는 실시예들을 설명하기 위한 것이며 본 발명을 제한하고자 하는 것은 아니다. 본 명세서에서, 단수형은 문구에서 특별히 언급하지 않는 한 복수형도 포함한다. 명세서에서 사용되는 "포함한다(comprises)" 및/또는 "포함하는(comprising)"은 언급된 구성요소 외에 하나 이상의 다른 구성요소의 존재 또는 추가를 배제하지 않는다. 명세서 전체에 걸쳐 동일한 도면 부호는 동일한 구성 요소를 지칭하며, "및/또는"은 언급된 구성요소들의 각각 및 하나 이상의 모든 조합을 포함한다. 비록 "제1", "제2" 등이 다양한 구성요소들을 서술하기 위해서 사용되나, 이들 구성요소들은 이들 용어에 의해 제한되지 않음은 물론이다. 이들 용어들은 단지 하나의 구성요소를 다른 구성요소와 구별하기 위하여 사용하는 것이다. 따라서, 이하에서 언급되는 제1 구성요소는 본 발명의 기술적 사상 내에서 제2 구성요소일 수도 있음은 물론이다.
- [0015] 다른 정의가 없다면, 본 명세서에서 사용되는 모든 용어(기술 및 과학적 용어를 포함)는 본 발명이 속하는 기술 분야의 통상의 기술자에게 공통적으로 이해될 수 있는 의미로 사용될 수 있을 것이다. 또한, 일반적으로 사용되는 사전에 정의되어 있는 용어들은 명백하게 특별히 정의되어 있지 않는 한 이상적으로 또는 과도하게 해석되지 않는다.
- [0016] 본 명세서에서, 전자 장치는 적어도 하나의 프로세서를 포함하는 무선 통신 기능을 포함하는 하드웨어 장치를 의미하는 것이고, 실시 예에 따라 해당 하드웨어 장치에서 동작하는 소프트웨어적 구성도 포괄하는 의미로서 이해될 수 있다. 예를 들어, 전자 장치는 스마트폰, 휴대폰, 핸드헬드 장치(hand-held device), 태블릿 PC, 데스크톱, 노트북 및 각 장치에서 구동되는 사용자 클라이언트 및 애플리케이션을 모두 포함하는 의미로서 이해될 수 있으며, 또한 이에 제한되는 것은 아니다.
- [0017] 이하, 첨부된 도면을 참조하여 본 발명의 실시예를 상세하게 설명한다.
- [0018] 본 명세서에서 설명되는 각 단계들은 단말에 의하여 수행되는 것으로 설명되나, 각 단계의 주체는 이에 제한되는 것은 아니며, 실시 예에 따라 각 단계들의 적어도 일부가 서로 다른 장치에서 수행될 수도 있다.
- [0019] 도 1은 본 발명의 일 실시예에 따른 아바타 생성 시스템(1)을 도시한 도면이며, 도 2는 본 발명의 일 실시예에 따른 아바타 생성 시스템(1)의 동작 방법을 도시한 순서도이다.
- [0020] 본 발명의 일 실시예에 따른 아바타 생성 시스템(1)은 서버(10), 복수의 전자 장치(20)를 포함한다. 서버(10)와 적어도 하나의 전자 장치(20) 및 각각의 전자 장치(20)는 서로 네트워크를 통해 통신할 수 있으며, 이때 각각의 전자 장치(20)는 클라이언트 장치(client device)로 동작할 수 있다.
- [0021] 서버(10)는 적어도 하나의 전자 장치(20)로부터 복수의 비디오 데이터(또는 복수의 비디오 프레임(frame))를 수신할 수 있다(S21). 각 비디오 데이터(또는 비디오 프레임)는 음성 데이터 및 이미지 데이터를 포함하며, 서버(10)는 수신된 비디오 데이터로부터 음성 데이터 및 이미지 데이터를 추출할 수 있다. 이렇게 추출된 이미지 데이터는 2차원 이미지 데이터일 수 있다.

- [0022] 서버(10)는 이미지 데이터에서 제1 랜드마크(landmark)를 검출하고(S22), 검출된 제1 랜드마크로부터 이미지 인코딩 값을 추출한다(S23). 보다 구체적으로, 서버(10)는 이미지 데이터에서 사람 및/또는 동물의 얼굴 영역을 검출한 후, 검출된 영역을 입력 이미지에서 기 결정된 개수의 특징점을 검출하도록 기 학습된 신경망에 입력하여 제1 랜드마크를 추출할 수 있다. 예를 들어, 서버(10)는 이미지 데이터를 입력 이미지에서 얼굴 영역을 인지하도록 기 학습된 CNN(Convolutional Neural Network) 기반의 신경망 등에 입력하거나, Haar-like Feature 알고리즘 등을 이용하여 이미지 데이터로부터 얼굴 영역을 추출할 수 있다. CNN은 특징 추출(Feature extraction) 영역과 이미지 분류(Classification) 영역으로 구분되며, 이미지를 분석하기 위한 패턴을 찾는 데 유용한 알고리즘이다. 서버(10)는, 보다 구체적인 예로서, ImageNet으로 학습된 CNN 기반의 AlexNet 신경망을 이용하여 이미지 데이터로부터 얼굴 영역을 추출할 수 있으나 이에 한정되는 것은 아니다.
- [0023] 또한, 제1 랜드마크는 얼굴 영역에 위치하는 특징점들의 집합으로서, 예를 들어, 이차원 벡터값의 집합($P = \{p_1, p_2, \dots, p_N\}$, N 은 정수)으로 표현될 수 있다. 서버(10)는, 예를 들어, 검출된 얼굴 영역을 기 학습된 CNN 기반의 신경망에 입력하여 제1 랜드마크로 변환할 수 있으며, 보다 구체적인 예로서, OpenCV의 Dlib, CLM-Framework, Google의 Mediapipe 등을 이용할 수 있으나 이에 한정되는 것은 아니다.
- [0024] 한편, 위 설명에서는 서버(10)가 이미지 데이터에서 얼굴 영역을 검출한 후에 얼굴 영역으로부터 제1 랜드마크를 검출하는 것으로 설명하였으나, 구현예에 따라, 서버(10)는 이미지 데이터를 기 학습된 CNN 기반의 신경망에 입력하여 제1 랜드마크를 검출할 수도 있다. 또한, 서버(10)는 전자 장치(20)로부터 얼굴 영역만을 포함하는 비디오 데이터를 수신할 수도 있다. 이 경우, 전자 장치(20)가 촬영된 비디오 데이터 또는 저장된 비디오 데이터로부터 얼굴 영역을 검출한 후에 서버(10)로 송신할 수 있다.
- [0025] 서버(10)는 제1 랜드마크를 기 학습된 MLP(Multi-Layer Perceptron) 기반의 신경망에 입력하여 이미지 인코딩 값으로 변환할 수 있다. 도 3은 본 발명의 일 실시예에 따른 MLP를 도시한 일예이다. 서버(10)는 제1 랜드마크(P)를 인코딩하도록 기 학습된 MLP에 입력한 후 MLP의 은닉층(Hidden Layer) 및 출력층(Output Layer)의 활성화 함수에 의해 이미지 인코딩 값(Y)을 획득할 수 있다. 한편, MLP 구조의 적합 능력은 은닉층의 개수 및 초기 가중 값에 영향을 받으며, 도 3에서는 두 개의 은닉층을 사용하는 것으로 도시하였으나 이에 한정되는 것은 아니다.
- [0026] 다시 도 2를 참조하면, 서버(10)는 음성 데이터로부터 콘텐츠 인코딩 값을 추출한다(S24). 서버(10)는 음성 데이터를 샘플링(sampling)하고, 샘플링된 값을 AutoEncoder 기반의 딥러닝 방식, DNN(Deep Neural Network) 기반의 딥러닝 방식 등을 이용하여 콘텐츠 인코딩 값을 추출할 수 있으나 이에 한정되는 것은 아니다. 이때, 콘텐츠 인코딩 값은 이미지 인코딩 값과 동일한 개수일 수 있으나 이에 한정되는 것은 아니다.
- [0027] 한편, 위 설명에서는 서버(10)가 이미지 인코딩 값을 추출한 이후에 콘텐츠 인코딩 값을 추출하는 것으로 기술하였으나, 이에 제한되는 것은 아니다. 서버(10)는 이미지 인코딩 값과 콘텐츠 인코딩 값을 동시에 추출할 수도 있으며, 콘텐츠 인코딩 값을 먼저 추출할 수도 있다.
- [0028] 이후, 서버(10)는 이미지 인코딩 값과 콘텐츠 인코딩 값에 기반하여 제2 랜드마크를 생성한다(S25). 서버(10)는 이미지 인코딩 값에 콘텐츠 인코딩 값을 가산, 감산, AND, OR, XOR 및 시프팅(Shifting) 중 적어도 하나의 논리 연산하여 새로운 제2 랜드마크로 변환할 수 있으며, 이미지 인코딩 값과 콘텐츠 인코딩 값을 적어도 하나의 논리 연산을 수행하도록 기 학습된 MLP 기반의 신경망에 입력하여 제2 랜드마크로 변환할 수도 있으나 이에 한정되는 것은 아니다. 한편, 제2 랜드마크는 2차원 벡터값일 수 있다.
- [0029] 서버(10)는 제1 랜드마크 및 제2 랜드마크를 이용하여 3차원의 포즈 정보, 기하 정보 및 표현 정보를 생성한다(S26). 서버(10)는 제1 랜드마크 및 제2 랜드마크를 3DMM(3D Morphable face Model) 기반의 신경망에 입력하여 복수의 파라미터(parameter)로 변환할 수 있다. 3DMM은 3D 스캐너나 깊이 카메라를 이용하여 얼굴의 3차원 정보를 직접적으로 이용하여 3차원 형태와 질감 간의 통계적 모델을 표현한 것으로, 본 발명의 일 실시예에서는 3DMM을 이용하여 2차원 랜드마크 값들로부터 3차원 포즈 파라미터(pose parameter), 기하 파라미터(geometry parameter) 및 표현 파라미터(expression parameter)를 획득한다. 보다 구체적으로, 서버(10)는 제1 랜드마크를 기 학습된 3DMM 기반의 신경망에 입력하여 포즈 파라미터로 변환하고, 제2 랜드마크를 기 학습된 3DMM 기반의 신경망에 입력하여 기하 파라미터 및 표현 파라미터로 변환할 수 있다. 이를 통해, 서버(10)는 사람(또는 동물)의 발화에 따른 입부분의 움직임뿐 아니라 얼굴의 전체적 상태를 포함하는 3차원 데이터를 획득할 수 있다.
- [0030] 이후, 서버(10)는 3차원 포즈 정보, 기하 정보 및 표현 정보를 기초로 2차원 평면으로 정사영된 아바타 이미지를 생성한다(S27). 도 4는 3차원 정보로부터 아바타 이미지를 생성하는 방법을 구체적으로 도시한 순서도이다.

도 4를 참조하면, 서버(10)는 3차원 포즈 정보, 기하 정보 및 표현 정보로부터 3차원 메쉬(mesh)를 생성하고 (S41), 3차원 메쉬의 특징점을 2차원 평면으로 정사영하여 제3 랜드마크를 추출한다(S42). 하기 수식 1은 2차원 평면으로 정사영된 제3 랜드마크(l_p)를 추출하는 방법을 도시한다.

[수식 1]

$$l_p = s_t * Pr * R_t * L_p^{3d} + t_t,$$

수식 1에서 (R_t , t_t , S_t)는 포즈 파라미터의 값이며, Pr 은 정사영 행렬 ($Pr = \{(1,0,0), (0,1,0)\}$) 이다. L_p^{3d} 는 3차원 메쉬의 특징점이다.

서버(10)는 제3 랜드마크를 기 학습된 cGAN(Conditional Generative Adversarial Network) 기반의 신경망에 입력하여 아바타 이미지를 생성한다(S43). 이때, 기 학습된 cGAN은 입력 데이터로부터 얼굴 이미지를 생성하는 조건부 확률이 추가되어 학습된 신경망일 수 있다. 서버(10)는 제3 랜드마크에 대한 전처리 작업을 수행할 수 있다. 예를 들어, 서버(10)는 입력 이미지로부터 경계 값을 인코딩하도록 기 학습된 신경망(예컨대, CNN 등)에 제3 랜드마크를 입력하여 경계 파라미터(edge parameter)로 변환하고, 경계 파라미터를 cGAN의 입력값으로 할 수도 있다.

서버(10)는 생성된 아바타 이미지 및 음성 데이터를 적어도 하나의 전자 장치(20)로 전송하여 전자 장치(20)에서 아바타 이미지가 음성 데이터와 함께 출력되도록 할 수 있다.

도 5는 본 발명의 일 실시예에 따라 도 2 내지 도 4의 동작을 서버(10)가 아닌 전자 장치(20)가 수행하는 아바타 생성 시스템(2)을 도시한 일 예이다. 즉, 도 2 내지 도 4에서는 서버(10)가 도 2 내지 도 4의 방법을 수행하는 것으로 기술하였으나, 본 발명의 구현예에 따라, 도 5에 도시된 바와 같이 하나의 전자 장치(20)가 상기 도 2 내지 도 4의 방법을 수행할 수 있다. 예를 들어, 하나의 전자 장치(20)는 촬영된 비디오 데이터 또는 저장된 비디오 데이터로부터 이미지 데이터 및 음성 데이터를 추출하고, 이미지 데이터로부터 검출된 제1 랜드마크(landmark)로부터 이미지 인코딩 값을 추출하고, 음성 데이터로부터 콘텐츠 인코딩 값을 추출하고, 이미지 인코딩 값과 콘텐츠 인코딩 값에 기반하여 제2 랜드마크를 추출하고, 제 1 랜드마크 및 제 2 랜드마크를 이용하여 3차원의 포즈 정보, 기하 정보 및 표현 정보를 생성하고, 3차원의 포즈 정보, 기하 정보 및 표현 정보를 기초로 2차원 평면으로 정사영된 아바타 이미지를 생성할 수 있다. 생성된 아바타 이미지는 네트워크로 연결된 다른 전자 장치들로 전송될 수 있다.

또한, 구현예에 따라, 도 2내지 도 4의 방법을 복수의 전자 장치들이 분산 수행하는 것도 가능하다는 것은 본 발명의 당업자에게 용이하게 이해될 수 있으므로 자세한 설명은 생략한다.

도 6은 본 발명의 일 실시예에 따른 아바타 생성 방법을 도시한 도면이다. 도 6의 실시예에서는 제1 랜드마크와 제2 랜드마크를 XOR 연산한 결과값을 기 학습된 3DMM 기반의 신경망에 입력하여 기하 파라미터 및 표현 파라미터를 생성하는 것으로 도시하였으나 이에 한정되는 것은 아니다.

도 7은 본 발명의 일 실시 예에 따른 서버(10)의 구성도이다.

프로세서(710)는 하나 이상의 코어(core, 미도시) 및 그래픽 처리부(미도시) 및/또는 다른 구성 요소와 신호를 송수신하는 연결 통로(예를 들어, 버스(bus) 등)를 포함할 수 있다.

일 실시예에 따른 프로세서(710)는 메모리(720)에 저장된 하나 이상의 인스트럭션을 실행함으로써, 도 1 내지 도 7과 관련하여 설명된 방법을 수행한다. 예를 들어, 프로세서(710)는 비디오 데이터로부터 이미지 데이터 및 음성 데이터를 추출하고, 이미지 데이터로부터 검출된 제1 랜드마크(landmark)로부터 이미지 인코딩 값을 추출하며, 상기 음성 데이터로부터 콘텐츠 인코딩 값을 추출하고, 이미지 인코딩 값과 콘텐츠 인코딩 값에 기반하여 제2 랜드마크를 추출하며, 제1 랜드마크 및 제2 랜드마크를 이용하여 3차원의 포즈 정보, 기하 정보 및 표현 정보를 생성한다. 구체적으로, 프로세서(710)는 제1 랜드마크를 기 학습된 3DMM(3D Morphable face Model) 기반의 신경망에 입력하여 포즈 파라미터로 변환하고, 제2 랜드마크를 기 학습된 3DMM 기반의 신경망에 입력하여 기하 파라미터 및 표현 파라미터로 변환할 수 있다.

프로세서(710)는 3차원의 포즈 정보, 기하 정보 및 표현 정보를 기초로 2차원 평면으로 정사영된 아바타 이미지

를 생성한다. 구체적으로, 프로세서(710)는 3차원 포즈 정보, 기하 정보 및 표현 정보로부터 3차원 메쉬(mesh)를 생성하고, 3차원 메쉬의 특징점을 2차원 평면으로 정사영하여 제3 랜드마크를 추출하며, 제3 랜드마크를 학습된 cGAN(Conditional Generative Adversarial Network) 기반의 신경망에 입력하여 상기 아바타 이미지를 생성할 수 있다.

[0043] 한편, 프로세서(710)는 프로세서(710) 내부에서 처리되는 신호(또는, 데이터)를 일시적 및/또는 영구적으로 저장하는 램(RAM: Random Access Memory, 미도시) 및 롬(ROM: Read-Only Memory, 미도시)을 더 포함할 수 있다. 또한, 프로세서(710)는 그래픽 처리부, 램 및 롬 중 적어도 하나를 포함하는 시스템 온 칩(SoC: system on chip) 형태로 구현될 수 있다.

[0044] 메모리(720)에는 프로세서(710)의 처리 및 제어를 위한 프로그램들(하나 이상의 인스트럭션들)을 저장할 수 있다. 메모리(720)에 저장된 프로그램들은 기능에 따라 복수 개의 모듈들로 구분될 수 있다.

[0045] 또한, 서버(10)는 상기의 구성들 이외에, 각 기능에서 처리되는 데이터 등을 표시하고 사용자 입력(예컨대, 터치 등)을 수신하는 디스플레이부(미도시), 다른 전자 장치 등과 통신하기 위한 적어도 하나의 구성요소를 포함하는 통신부(미도시), 사용자의 음성을 입력받는 마이크(미도시) 등을 더 포함할 수 있다. 그러나, 본 발명의 서버(10)가 더 많은 구성요소들을 포함하여 구현될 수 있음은 본 발명의 기술분야의 당업자가 용이하게 이해할 수 있을 것이다.

[0046] 한편, 도 7의 구성요소들은 전자 장치(20)의 구성들에 대응될 수 있다. 즉, 전자 장치(20)는 하나 이상의 코어 및 그래픽 처리부를 포함하는 프로세서 및 메모리를 포함하며, 프로세서는 메모리에 저장된 하나 이상의 인스트럭션을 실행함으로써 도 1 내지 도 6과 관련하여 설명된 방법을 수행할 수 있다.

[0047] 한편, 본 발명의 실시예와 관련하여 설명된 방법 또는 알고리즘의 단계들은 하드웨어로 직접 구현되거나, 하드웨어에 의해 실행되는 소프트웨어 모듈로 구현되거나, 또는 이들의 결합에 의해 구현될 수 있다. 소프트웨어 모듈들은 RAM(Random Access Memory), ROM(Read Only Memory), EPROM(Erasable Programmable ROM), EEPROM(Electrically Erasable Programmable ROM), 플래시 메모리(Flash Memory), 하드 디스크, 착탈형 디스크, CD-ROM, 또는 본 발명이 속하는 기술 분야에서 잘 알려진 임의의 형태의 컴퓨터 판독가능 기록매체에 상주할 수도 있다.

[0048] 본 발명의 구성 요소들은 하드웨어인 컴퓨터와 결합되어 실행되기 위해 프로그램(또는 애플리케이션)으로 구현되어 매체에 저장될 수 있다. 예를 들어, 본 발명의 일 실시예는 컴퓨터에서 독출가능한 기록매체에 저장된 컴퓨터프로그램으로서, 비디오 데이터로부터 이미지 데이터 및 음성 데이터를 추출하는 기능, 이미지 데이터로부터 검출된 제1 랜드마크(landmark)로부터 이미지 인코딩 값을 추출하는 기능, 음성 데이터로부터 콘텐츠 인코딩 값을 추출하는 단계, 이미지 인코딩 값과 콘텐츠 인코딩 값에 기반하여 제2 랜드마크를 추출하는 기능, 제1 랜드마크 및 제2 랜드마크를 이용하여 3차원의 포즈 정보, 기하 정보 및 표현 정보를 생성하는 기능, 및 3차원의 포즈 정보, 기하 정보 및 표현 정보를 기초로 2차원 평면으로 정사영된 아바타 이미지를 생성하는 기능을 수행할 수 있다.

[0049] 또한, 본 발명의 구성 요소들은 소프트웨어 프로그래밍 또는 소프트웨어 요소들로 실행될 수 있으며, 이와 유사하게, 실시 예는 데이터 구조, 프로세스들, 루틴들 또는 다른 프로그래밍 구성들의 조합으로 구현되는 다양한 알고리즘을 포함하여, C, C++, 자바(Java), 어셈블러(assembler), 파이썬(Python) 등과 같은 프로그래밍 또는 스크립팅 언어로 구현될 수 있다. 기능적인 측면들은 하나 이상의 프로세서들에서 실행되는 알고리즘으로 구현될 수 있다.

[0050] 이상의 설명은 본 발명의 기술 사상을 예시적으로 설명한 것에 불과한 것으로서, 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자라면 본 발명의 본질적인 특성에서 벗어나지 않는 범위에서 다양한 수정 및 변형이 가능할 것이다. 또한, 본 발명에 개시된 실시 예들은 본 발명의 기술 사상을 한정하기 위한 것이 아니라 설명하기 위한 것이고, 이러한 실시 예에 의하여 본 발명의 기술 사상의 범위가 한정되는 것은 아니다. 따라서, 본 발명의 보호 범위는 아래의 청구범위에 의하여 해석되어야 하며, 그와 동등한 범위 내에 있는 모든 기술 사상은 본 발명의 권리범위에 포함되는 것으로 해석되어야 할 것이다.

부호의 설명

[0051] 1, 2: 아바타 생성 시스템
10: 서버

20: 전자 장치

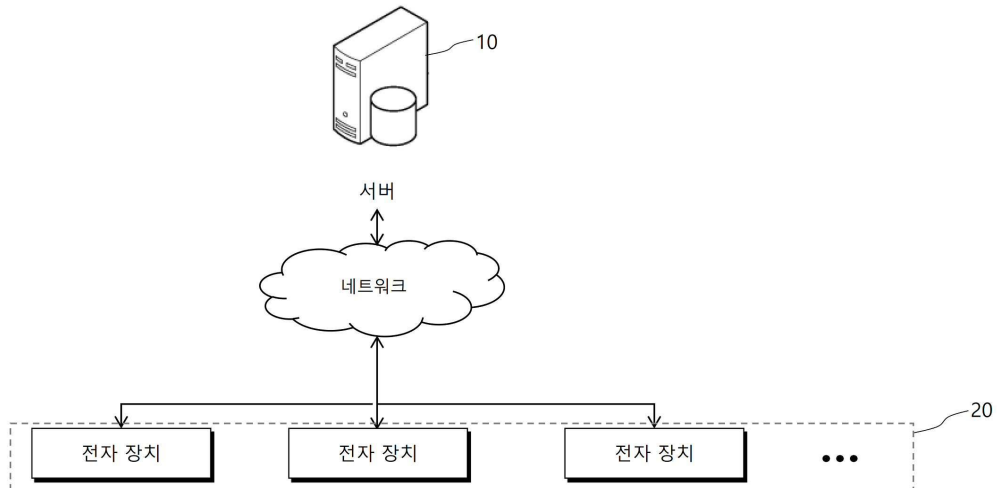
710: 프로세서

720: 메모리

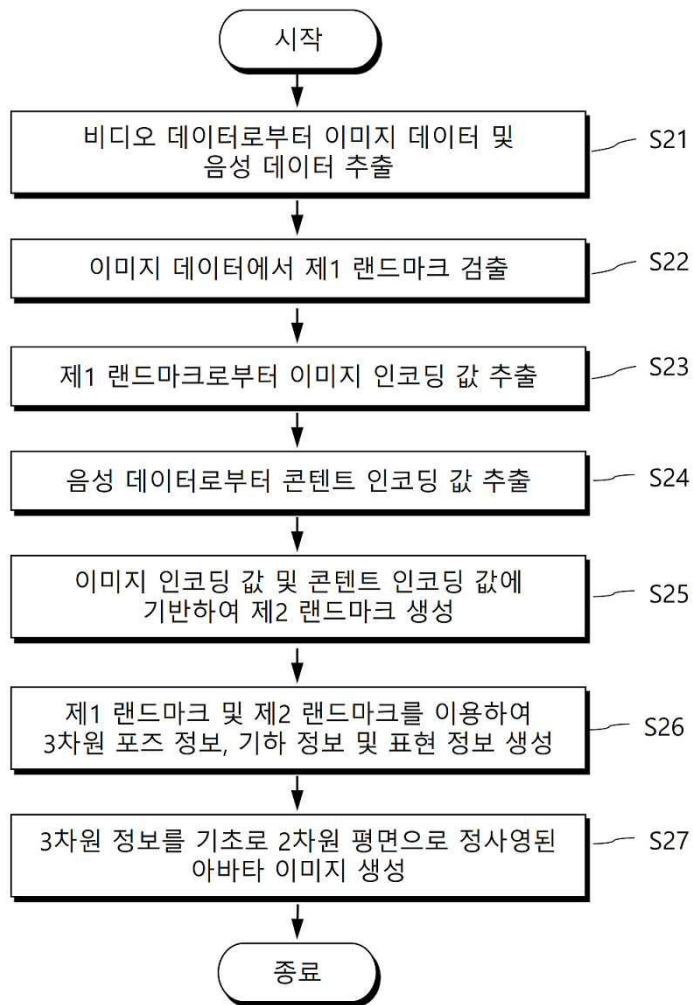
도면

도면1

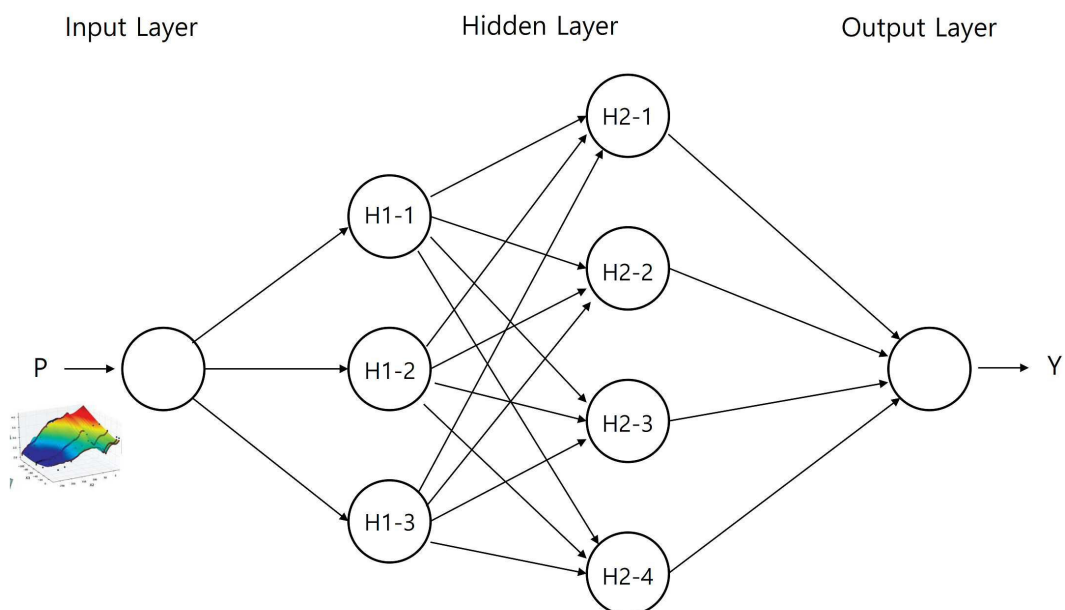
1



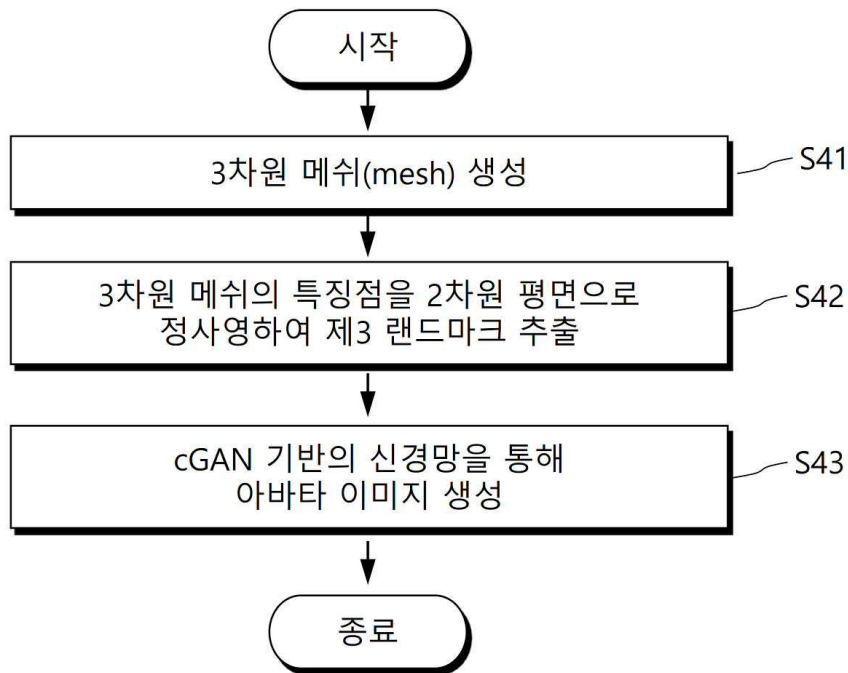
도면2



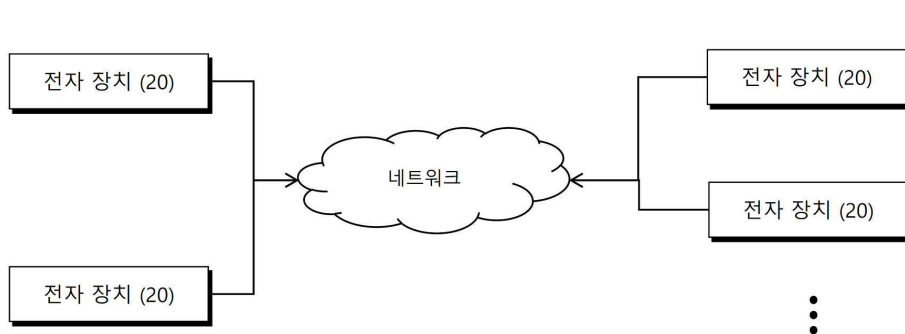
도면3



도면4

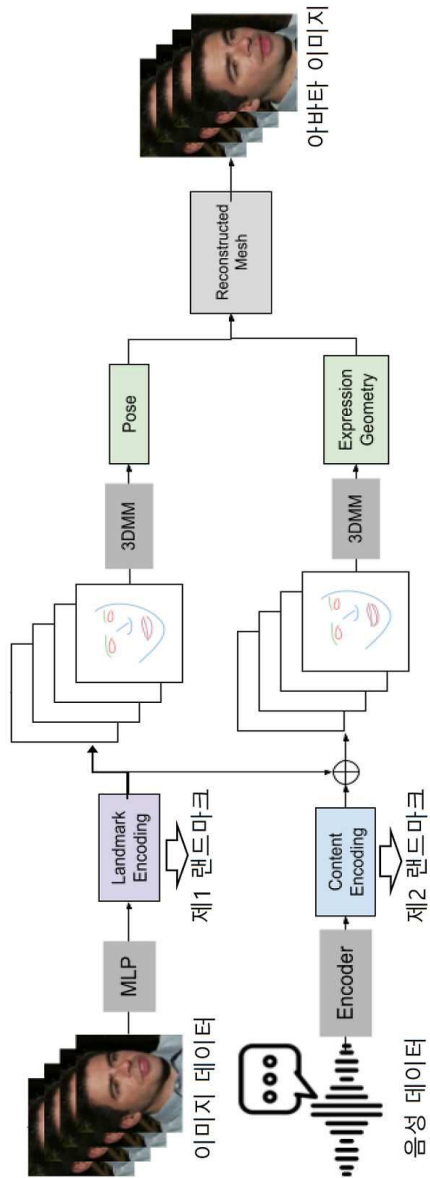


도면5



2

도면6



도면7

