



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2020년12월31일
(11) 등록번호 10-2196686
(24) 등록일자 2020년12월23일

(51) 국제특허분류(Int. Cl.)
G06F 21/32 (2013.01) G06K 9/00 (2006.01)
G10L 15/08 (2006.01) H04L 9/32 (2006.01)
(52) CPC특허분류
G06F 21/32 (2013.01)
G06K 9/00221 (2013.01)
(21) 출원번호 10-2018-7036914
(22) 출원일자(국제) 2017년04월12일
심사청구일자 2019년05월23일
(85) 번역문제출일자 2018년12월19일
(65) 공개번호 10-2019-0009361
(43) 공개일자 2019년01월28일
(86) 국제출원번호 PCT/CN2017/080196
(87) 국제공개번호 WO 2017/198014
국제공개일자 2017년11월23일
(30) 우선권주장
201610340549.4 2016년05월19일 중국(CN)
(56) 선행기술조사문헌
US20130267204 A1

(73) 특허권자
어드밴스드 뉴 테크놀로지스 씨오., 엘티디.
케이만 군도, 그랜드 케이만 케이와이1-9008, 조지 타운, 27 하스피탈 로드, 케이만 코포레이트 센터
(72) 발명자
리 펑
중국 저지앙 311121 항저우 유 향 디스트릭트 넘버 969 웨스트 웬 이 로드 빌딩 3 알리바바 그룹 리갈 디파트먼트 5층
선 이펑
중국 저지앙 311121 항저우 유 향 디스트릭트 넘버 969 웨스트 웬 이 로드 빌딩 3 알리바바 그룹 리갈 디파트먼트 5층
(뒷면에 계속)
(74) 대리인
김태홍, 김진희

전체 청구항 수 : 총 20 항

심사관 : 문남두

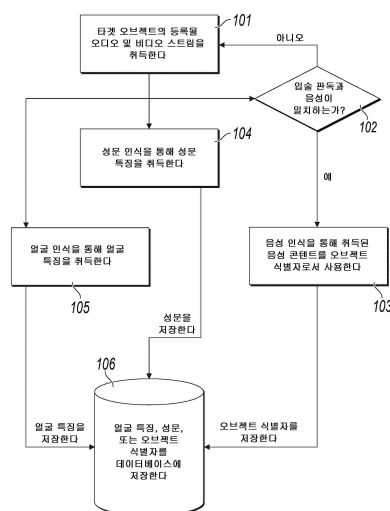
(54) 발명의 명칭 신원 인증 방법 및 장치

(57) 요약

본 출원은 신원 인증 방법 및 장치를 제공한다. 상기 방법은, 인증될 타겟 오브젝트에 의해 생성된 수집된 오디오 및 비디오 스트림을 취득하는 단계; 상기 오디오 및 비디오 스트림에서의 입술 판독과 음성이 일치하는지 여부를 결정하고, 상기 입술 판독과 상기 음성이 일치하면, 상기 오디오 및 비디오 스트림에서의 오디오 스트림에

(뒷면에 계속)

대표도 - 도1



대하여 음성 인식을 수행함으로써 취득된 음성 콘텐츠를 상기 타겟 오브젝트의 오브젝트 식별자로서 사용하는 단계; 미리 저장된 오브젝트 등록 정보가 상기 오브젝트 식별자를 포함하면, 상기 오브젝트 등록 정보로부터 상기 오브젝트 식별자에 대응하는 모델 생리 특징을 취득하는 단계; 상기 타겟 오브젝트의 생리 특징을 취득하기 위해 상기 오디오 및 비디오 스트림에 대하여 생리 인식을 수행하는 단계; 및 상기 타겟 오브젝트의 상기 생리 특징을 상기 모델 생리 특징과 비교하여 비교 결과를 취득하고, 상기 비교 결과가 인증 조건을 충족시키면, 상기 타겟 오브젝트가 인증된 것으로 결정하는 단계를 포함한다. 본 출원은 신원 인증의 효율 및 신뢰성을 향상시킨다.

(52) CPC특허분류

G06K 9/00335 (2013.01)

G10L 15/08 (2013.01)

H04L 9/3231 (2013.01)

(72) 발명자

시에 용시양

중국 저지양 311121 항저우 유 향 디스트릭트 넘버
969 웨스트 웨 이 로드 빌딩 3 알리바바 그룹 리갈
디파트먼트 5층

리 리양

중국 저지양 311121 항저우 유 향 디스트릭트 넘버
969 웨스트 웨 이 로드 빌딩 3 알리바바 그룹 리갈
디파트먼트 5층

명세서

청구범위

청구항 1

컴퓨터 구현 방법(computer-implemented method)으로서,

인증될 사용자의 오디오/비디오 스트림을 취득하는 단계;

상기 오디오/비디오 스트림 내의 상기 사용자의 음성이 상기 오디오/비디오 스트림 내의 상기 사용자의 입술과 매칭되는 것을 결정하는 단계;

상기 오디오/비디오 스트림 내의 상기 사용자의 음성이 상기 오디오/비디오 스트림 내의 상기 사용자의 입술과 매칭된다는 결정에 응답하여, 상기 오디오/비디오 스트림에 대한 자동화된 음성 인식을 수행하는 것에 기초하여, 상기 사용자에 대한 사용자 식별자를 결정하는 단계;

상기 오디오/비디오 스트림에 대한 자동화된 생리 특징 추출(automated physiological feature extraction)을 수행하는 것에 기초하여, 상기 사용자의 생리 특징을 결정하는 단계;

저장된 오브젝트 등록 정보로부터, 상기 결정된 사용자 식별자에 대응하는 저장된 모델 생리 특징을 취득하는 단계;

상기 오디오/비디오 스트림에 대한 자동화된 생리 특징 추출을 수행하는 것에 기초하여 결정된 상기 사용자의 생리 특징을, 상기 저장된 모델 생리 특징과 비교하는 것에 기초하여, 비교 결과를 생성하는 단계; 및

상기 비교 결과가 인증 조건을 만족한다는 결정에 응답하여, 상기 사용자가 인증되었다고 결정하는 단계를 포함하는, 컴퓨터 구현 방법.

청구항 2

제1항에 있어서,

상기 사용자의 생리 특징은 상기 사용자의 얼굴 특징을 포함하는 것인, 컴퓨터 구현 방법.

청구항 3

제1항에 있어서,

상기 비교 결과는 유사성 스코어(similarity score)를 포함하는 것인, 컴퓨터 구현 방법.

청구항 4

제1항에 있어서,

상기 비교 결과가 인증 조건을 만족한다고 결정하는 것은, 유사성 스코어가 문턱 스코어(threshold score)를 초과한다고 결정하는 것을 포함하는 것인, 컴퓨터 구현 방법.

청구항 5

제1항에 있어서,

상기 오디오/비디오 스트림 내의 상기 사용자의 음성이 상기 사용자의 입술과 매칭되는 것을 결정하는 단계는,

특정 시간 포인트(particular point in time)에서 상기 오디오/비디오 스트림의 비디오 이미지 내의 입술 판독 음절(lip reading syllable)을 결정하는 단계;

상기 특정 시간 포인트에서의 상기 오디오/비디오 스트림의 오디오 내의 음성 음절(voice syllable)을 결정하는 단계; 및

상기 입술 판독 음절과 상기 음성 음절이 매칭되는 것을 결정하는 단계

를 포함하는 것인, 컴퓨터 구현 방법.

청구항 6

제1항에 있어서,

상기 오브젝트 등록 정보에 상기 모델 생리 특징을 저장하는 단계를 포함하는, 컴퓨터 구현 방법.

청구항 7

제1항에 있어서,

인증을 위해 상기 사용자로부터 요청을 수신하는 단계를 포함하고, 상기 사용자의 오디오/비디오 스트림은 상기 요청에 응답하여 취득되는 것인, 컴퓨터 구현 방법.

청구항 8

동작들을 수행하기 위해 컴퓨터 시스템에 의해 실행 가능한 하나 이상의 명령어가 저장된 비일시적 컴퓨터 판독 가능 매체로서, 상기 동작들은,

인증될 사용자의 오디오/비디오 스트림을 취득하는 동작;

상기 오디오/비디오 스트림 내의 상기 사용자의 음성이 상기 오디오/비디오 스트림 내의 상기 사용자의 입술과 매칭되는 것을 결정하는 동작;

상기 오디오/비디오 스트림 내의 상기 사용자의 음성이 상기 오디오/비디오 스트림 내의 상기 사용자의 입술과 매칭된다는 결정에 응답하여, 상기 오디오/비디오 스트림에 대한 자동화된 음성 인식을 수행하는 것에 기초하여, 상기 사용자에 대한 사용자 식별자를 결정하는 동작;

상기 오디오/비디오 스트림에 대한 자동화된 생리 특징 추출을 수행하는 것에 기초하여, 상기 사용자의 생리 특징을 결정하는 동작;

저장된 오브젝트 등록 정보로부터, 상기 결정된 사용자 식별자에 대응하는 저장된 모델 생리 특징을 취득하는 동작;

상기 오디오/비디오 스트림에 대한 자동화된 생리 특징 추출을 수행하는 것에 기초하여 결정된 상기 사용자의 생리 특징을, 상기 저장된 모델 생리 특징과 비교하는 것에 기초하여, 비교 결과를 생성하는 동작; 및

상기 비교 결과가 인증 조건을 만족한다는 결정에 응답하여, 상기 사용자가 인증되었다고 결정하는 동작을 포함하는 것인, 비일시적 컴퓨터 판독 가능 매체.

청구항 9

제8항에 있어서,

상기 사용자의 생리 특징은 상기 사용자의 얼굴 특징을 포함하는 것인, 비일시적 컴퓨터 판독 가능 매체.

청구항 10

제8항에 있어서,

상기 비교 결과는 유사성 스코어를 포함하는 것인, 비일시적 컴퓨터 판독 가능 매체.

청구항 11

제8항에 있어서,

상기 비교 결과가 인증 조건을 만족한다고 결정하는 것은, 유사성 스코어가 문턱 스코어를 초과한다고 결정하는 것을 포함하는 것인, 비일시적 컴퓨터 판독 가능 매체.

청구항 12

제8항에 있어서,

상기 오디오/비디오 스트림 내의 상기 사용자의 음성이 상기 사용자의 입술과 매칭되는 것을 결정하는 동작은,
특정 시간 포인트에서 상기 오디오/비디오 스트림의 비디오 이미지 내의 입술 판독 음절을 결정하는 동작;

상기 특정 시간 포인트에서의 상기 오디오/비디오 스트림의 오디오 내의 음성 음절을 결정하는 동작;
및

상기 입술 판독 음절과 상기 음성 음절이 매칭되는 것을 결정하는 동작
을 포함하는 것인, 비밀시적 컴퓨터 판독 가능 매체.

청구항 13

제8항에 있어서,

상기 동작들은 상기 오브젝트 등록 정보에 상기 모델 생리 특징을 저장하는 동작을 포함하는 것인, 비밀시적 컴퓨터 판독 가능 매체.

청구항 14

제8항에 있어서,

상기 동작들은 인증을 위해 상기 사용자로부터 요청을 수신하는 동작을 포함하고, 상기 사용자의 오디오/비디오 스트림은 상기 요청에 응답하여 취득되는 것인, 비밀시적 컴퓨터 판독 가능 매체.

청구항 15

컴퓨터 구현 시스템(computer-implemented system)으로서,

하나 이상의 컴퓨터; 및

상기 하나 이상의 컴퓨터에 상호동작 가능하게 연결되고, 하나 이상의 명령어가 저장된 유형의(tangible) 비밀시적 기계 판독 가능 매체를 구비한 하나 이상의 컴퓨터 메모리 디바이스

를 포함하고,

상기 하나 이상의 명령어는, 상기 하나 이상의 컴퓨터에 의해 실행될 때, 하나 이상의 동작을 수행하고,

상기 하나 이상의 동작은,

인증될 사용자의 오디오/비디오 스트림을 취득하는 동작;

상기 오디오/비디오 스트림 내의 상기 사용자의 음성이 상기 오디오/비디오 스트림 내의 상기 사용자의 입술과 매칭되는 것을 결정하는 동작;

상기 오디오/비디오 스트림 내의 상기 사용자의 음성이 상기 오디오/비디오 스트림 내의 상기 사용자의 입술과 매칭된다는 결정에 응답하여, 상기 오디오/비디오 스트림에 대한 자동화된 음성 인식을 수행하는 것에 기초하여, 상기 사용자에게 대한 사용자 식별자를 결정하는 동작;

상기 오디오/비디오 스트림에 대한 자동화된 생리 특징 추출을 수행하는 것에 기초하여, 상기 사용자의 생리 특징을 결정하는 동작;

저장된 오브젝트 등록 정보로부터, 상기 결정된 사용자 식별자에 대응하는 저장된 모델 생리 특징을 취득하는 동작;

상기 오디오/비디오 스트림에 대한 자동화된 생리 특징 추출을 수행하는 것에 기초하여 결정된 상기 사용자의 생리 특징을, 상기 저장된 모델 생리 특징과 비교하는 것에 기초하여, 비교 결과를 생성하는 동작; 및

상기 비교 결과가 인증 조건을 만족한다는 결정에 응답하여, 상기 사용자가 인증되었다고 결정하는 동작

을 포함하는 것인, 컴퓨터 구현 시스템.

청구항 16

제15항에 있어서,

상기 사용자의 생리 특징은 상기 사용자의 얼굴 특징을 포함하는 것인, 컴퓨터 구현 시스템.

청구항 17

제15항에 있어서,

상기 비교 결과는 유사성 스코어를 포함하는 것인, 컴퓨터 구현 시스템.

청구항 18

제15항에 있어서,

상기 비교 결과가 인증 조건을 만족한다고 결정하는 것은, 유사성 스코어가 문턱 스코어를 초과한다고 결정하는 것을 포함하는 것인, 컴퓨터 구현 시스템.

청구항 19

제15항에 있어서,

상기 오디오/비디오 스트림 내의 상기 사용자의 음성이 상기 사용자의 입술과 매칭되는 것을 결정하는 동작은,

특정 시간 포인트에서 상기 오디오/비디오 스트림의 비디오 이미지 내의 입술 판독 음절을 결정하는 동작;

상기 특정 시간 포인트에서의 상기 오디오/비디오 스트림의 오디오 내의 음성 음절을 결정하는 동작;

및

상기 입술 판독 음절과 상기 음성 음절이 매칭되는 것을 결정하는 동작

청구항 20

제15항에 있어서,

상기 하나 이상의 동작은 상기 오브젝트 등록 정보에 상기 모델 생리 특징을 저장하는 동작을 포함하는 것인, 컴퓨터 구현 시스템.

발명의 설명

기술 분야

[0001] 본 출원은 네트워크 기술에 관한 것이고, 특히 신원 인증 방법 및 장치에 관한 것이다.

배경 기술

[0002] 인터넷 기술의 발전으로 인해, 네트워크 기반 서비스가 사람들에게 의해 광범위하게 사용된다. 예를 들어, 사람들은 메일박스를 사용하거나 온라인으로 쇼핑하거나 온라인으로 작업하여 전자 메일을 보내고 받을 수 있다. 일부 애플리케이션에는 높은 보안 요구사항이 있으며 사용자의 신원이 인증될 필요가 있다. 예를 들어, 온라인 쇼핑을 위한 지불이 인증되기 전에 사용자의 신원이 인증될 필요가 있고; 또는 사용자의 신원이 인증된 후에 비교적 높은 보안 요구사항을 가진 애플리케이션에 사용자가 로그인할 수 있다. 관련 기술에서, 얼굴 인식 및 성문(voiceprint) 인식과 같은 인터넷에 의해 사용되는 여러 가지 신원 인증 방법이 있습니다. 그러나, 이러한 일반적으로 사용되는 인증 방법은 비교적 복잡하다. 예를 들어, 사용자는 사용자의 ID를 입력하고 음성 인식을 위해 음성을 검증할 필요가 있다. 또한 기존의 인증 방법은 비교적 신뢰성이 낮다. 예를 들어, 공격자는 아날로그 비디오 또는 레코딩(recording)을 사용하여 얼굴 인식을 할 수 있다. 예를 들어, 얼굴과 성문을 참조하여 2가지 결합된 인증 방법에 기초하여 검증이 수행 되더라도, 이들 인증 방법은 비교적 서로 독립적이므로, 공격자는 인증 프로세스를 깰 수 있다. 전술한 신원 인증 방법의 단점은 애플리케이션 보안에 위협을 초래할 수 있다.

발명의 내용

- [0003] 따라서, 본 출원은 신원 인증의 효율성 및 신뢰성을 향상시키기 위한 신원 인증 방법 및 장치를 제공한다.
- [0004] 특히, 본 출원은 후술하는 기술 솔루션을 사용함으로써 구현된다.
- [0005] 제1 양태에 따르면, 신원 인증 방법이 제공되고, 상기 신원 인증 방법은, 수집된 오디오 및 비디오 스트림 - 상기 오디오 및 비디오 스트림은 인증될 타겟 오브젝트에 의해 생성됨 - 을 취득하는 단계; 상기 오디오 및 비디오 스트림의 입술 판독(lip reading)과 음성이 일치하는지 여부를 결정하고, 상기 입술 판독 및 상기 음성이 일치하는 경우, 상기 오디오 및 비디오 스트림의 오디오 스트림에 대하여 음성 인식을 수행함으로써 취득된 음성 콘텐츠를 상기 타겟 오브젝트의 오브젝트 식별자로서 사용하는 단계; 미리 저장된 오브젝트 등록 정보가 상기 오브젝트 식별자를 포함하면, 상기 상기 오브젝트 등록 정보로부터 상기 오브젝트 식별자에 대응하는 모델 생리 특징(model physiological feature)을 취득하는 단계; 상기 타겟 오브젝트의 생리 특징(physiological feature)을 취득하기 위해 상기 오디오 및 비디오 스트림에 대하여 생리 인식을 수행하는 단계; 및 상기 타겟 오브젝트의 생리 특징을 상기 모델 생리 특징과 비교하여 비교 결과를 취득하고, 상기 비교 결과가 인증 조건을 만족하면, 상기 타겟 오브젝트가 인증된 것으로 결정하는 단계를 포함한다.
- [0006] 제2 양태에 따르면, 신원 인증 장치가 제공되고, 상기 신원 인증 장치는, 수집된 오디오 및 비디오 스트림 - 상기 오디오 및 비디오 스트림은 인증될 타겟 오브젝트에 의해 생성됨 - 을 취득하도록 구성된 정보 취득 모듈; 상기 오디오 및 비디오 스트림의 입술 판독(lip reading)과 음성이 일치하는지 여부를 결정하고, 상기 입술 판독 및 상기 음성이 일치하는 경우, 상기 오디오 및 비디오 스트림에서의 오디오 스트림에 대하여 음성 인식을 수행함으로써 취득된 음성 콘텐츠를 상기 타겟 오브젝트의 오브젝트 식별자로서 사용하도록 구성된 식별자 결정 모듈; 미리 저장된 오브젝트 등록 정보가 상기 오브젝트 식별자를 포함하면, 상기 상기 오브젝트 등록 정보로부터 상기 오브젝트 식별자에 대응하는 모델 생리 특징(model physiological feature)을 취득하도록 구성된 정보 관리 모듈; 상기 타겟 오브젝트의 생리 특징을 취득하기 위해 상기 오디오 및 비디오 스트림에 대하여 생리 인식을 수행하도록 구성된 특징 인식 모듈; 및 상기 타겟 오브젝트의 생리 특징을 상기 모델 생리 특징과 비교하여 비교 결과를 취득하고, 상기 비교 결과가 인증 조건을 만족하면, 상기 타겟 오브젝트가 인증된 것으로 결정하도록 구성된 인증 프로세싱 모듈을 포함한다.
- [0007] 본 출원에서 제공된 신원 인증 방법 및 장치에 따르면, 사용자 식별자는 사용자의 인증 동안에 오디오 및 비디오 스트림의 인식을 통해 취득되고, 얼굴 특징 및 성문 특징은 동일한 오디오 및 비디오 스트림을 사용함으로써 검증될 수도 있다. 이것은, 사용자의 작업을 단순화하고, 인증 효율을 높이며, 일대일 인증 모델을 유지 관리하고, 인식 정확도를 보장한다. 또한, 이 방법에서는 입술 판독과 음성의 일치가 결정되어, 타겟 오브젝트가 공격자의 가짜 비디오 레코드가 아닌 살아있는 오브젝트임을 보장함으로써, 인증 보안 및 안정성을 향상시킨다.

도면의 간단한 설명

- [0008] 도 1은 본 출원의 예시적 구현예에 따른 신원 등록 절차를 예시한다.
- 도 2는 본 출원의 예시적 구현예에 따른 입술 판독과 음성 사이의 일치를 결정하는 절차를 예시한다.
- 도 3은 본 출원의 예시적 구현예에 따른 얼굴 특징 인식 절차를 예시한다.
- 도 4는 본 출원의 예시적 구현예에 따른 성문 특징 인식 절차를 예시한다.
- 도 5는 본 출원의 예시적 구현예에 따른 신원 인증 절차를 예시한다.
- 도 6은 본 출원의 예시적 구현예에 따른 신원 인증 장치의 구조적 다이어그램이다.
- 도 7은 본 출원의 예시적 구현예에 따른 신원 인증 장치의 구조적 다이어그램이다.

발명을 실시하기 위한 구체적인 내용

- [0009] 여기에 예시적 구현예가 상세히 설명되고, 예시적 구현예의 실시예가 첨부 도면에서 제시된다. 첨부 도면과 후속 설명이 관련될 때, 다르게 언급하지 않으면, 상이한 첨부 도면에서의 동일한 도면부호는 동일 또는 유사한 엘리먼트를 나타낸다. 이하의 예시적 구현예에서 설명되는 구현예는 본 출원에 따른 모든 구현예를 나타내지 않는다. 대신에, 이것들은 첨부된 청구 범위에서 상세히 설명되는 본 출원의 일부 양태와 일치하는 장치 및 방법의 단지 예일 뿐이다.

- [0010] 본 출원의 구현예에서 제공되는 신원 인증 방법은 인터넷 신원 인증에 적용될 수 있다. 예를 들어, 애플리케이션의 보안을 확보하기 위해, 사용자는 상기 방법을 사용함으로써 인증된 후에 네트워크 애플리케이션에 로그인할 수 있다.
- [0011] 다음은 비교적 높은 보안 요구사항을 가진 애플리케이션을 실시예로 사용한다. 예를 들어 스마트 폰 또는 지능형 태블릿과 같은 사용자의 지능형 디바이스에서 애플리케이션이 실행될 수 있다고 가정한다. 지능형 디바이스에서의 애플리케이션에 사용자가 로그인할 필요가 있는 경우에 지능형 디바이스에서의 카메라 및 마이크를 사용함으로써 오디오 및 비디오 스트림이 수집될 수 있다. 예를 들어, 사용자는 모바일 폰의 카메라 및 마이크를 사용자의 애플리케이션 ID를 판독할 수 있다. 애플리케이션 ID는 애플리케이션에서의 사용자에게 의해 등록된 계정 번호 "123456"이 될 수 있다. 사용자가 애플리케이션 ID를 판독한 후에, 모바일 폰은 비디오 이미지 및 판독 음성을 포함하는 사용자의 오디오 및 비디오 스트림을 수집할 수 있다.
- [0012] 본 출원의 이러한 구현예에서의 신원 인증 방법에 따르면, 수집된 오디오 및 비디오 스트림이 프로세싱될 수 있다. 인증 전에, 사용자는 나중에 신원 인증을 수행하기 위한 신원 등록 절차를 수행할 필요가 있다. 등록 절차는 또한 수집된 오디오 및 비디오 스트림에 기초하여 수행된다. 다음은 신원 등록 절차 및 신원 인증 절차를 개별적으로 설명한다. 또한 이 구현예는 실제 애플리케이션에서 신원 등록 또는 인증 프로세싱을 수행하는 디바이스에 대하여 제한을 부과하지 않는다. 예를 들어, 사용자의 오디오 및 비디오 스트림이 수집된 후, 스마트 폰은 오디오 및 비디오 스트림을 애플리케이션의 백엔드 서버(backend server)로 전송하거나, 프로세싱의 일부가 스마트 폰의 클라이언트 소프트웨어 측에서 수행될 수 있으며, 프로세싱의 다른 부분은 서버 측에서 수행될 수 있고; 또는 다른 방법이 사용될 수 있다.
- [0013] [신원 등록]
- [0014] 본 구현예의 방법에서, 사용자가 신원 등록을 수행할 때, 2가지 타입의 정보가 포함될 수 있다. 한가지 타입의 정보는 오브젝트 식별자이다. 예를 들어, 사용자가 애플리케이션에 로그인하는 경우, 이 사용자를 타겟 오브젝트라고 한다. 사용자가 애플리케이션에 등록하는 경우, 사용자를 다른 사용자와 구별하기 위해 사용되는 정보는, 예를 들어 애플리케이션에서의 사용자의 계정 번호 123456이 될 수 있는 오브젝트 식별자이고, 계정 번호는 타겟 오브젝트의 오브젝트 식별자이다. 다른 타입의 정보는 예를 들어 사용자의 성문 특징 또는 사용자의 얼굴 특징인 사용자를 고유하게 식별할 수 있는 생리 정보이다. 통상적으로, 상이한 사람들의 성문 및 얼굴은 상이하고 각 사용자를 식별하는 생리 정보는 모델 생리 특징이라고 할 수 있다.
- [0015] 맵핑 관계는 두 가지 타입의 정보, 즉 오브젝트 식별자와 모델 생리 특징 사이에 설정되며 맵핑 관계가 저장된다. 대응하게 저장된 타겟 오브젝트의 오브젝트 식별자 및 모델 생리 특징을 "오브젝트 등록 정보"라고 할 수 있다. 예를 들어, 사용자 Xiao Zhang은 사용자의 오브젝트 등록 정보를 "123456-모델 생리 특징 A"로 저장할 수 있다. 사용자의 더 정확한 식별을 위해, 본 실시예에서 사용되는 모델 생리 특징에 포함된 생리 정보는 예를 들어 얼굴 및 성문과 같은 적어도 2가지 타입의 생리 정보가 될 수 있다.
- [0016] 도 1은 예시적 신원 등록 절차를 예시한다. 상기 절차는 하기의 프로세싱을 포함한다:
- [0017] 단계 101: 타겟 오브젝트의 등록될 오디오 및 비디오 스트림을 취득한다.
- [0018] 예를 들어, 사용자는 애플리케이션에 등록하고, 사용자는 애플리케이션의 계정 번호 "123456"을 모바일 폰과 같은 사용자의 지능형 디바이스로 판독할 수 있다. 본 실시예에서, 등록을 수행하는 사용자는 타겟 오브젝트가 될 수 있고, 지능형 디바이스의 카메라 및 마이크는 사용자가 계정 번호를 판독할 때 생성되는 오디오 및 비디오 스트림을 수집할 수 있다. 등록 중에 수집된 오디오 및 비디오 스트림은 오디오 스트림 및 비디오 스트림을 포함하는 등록될 오디오 및 비디오 스트림이 될 수 있다. 오디오 스트림은 사용자의 음성을 판독하고, 비디오 스트림은 사용자의 판독 비디오 이미지이다.
- [0019] 현 단계에서 오디오 및 비디오 스트림이 취득된 후에, 사용자의 등록을 완료하기 위해 다음의 3개 양태의 프로세싱이 수행될 수 있다. 자세한 내용은 도 1을 참조할 수 있다.
- [0020] 제1 양태에서 프로세싱은 다음과 같다: 단계 102: 등록될 오디오 및 비디오 스트림에서의 입술 판독 및 음성이 일치하는지 여부를 결정한다.
- [0021] 여기서의 일치는 입술 움직임과 음성으로 표시된 움직임 간의 매핑 관계를 나타낸다. 예를 들어, "오늘 날씨는 맑다"라는 음성이 저속으로 생성되지만, "오늘 날씨는 맑다"에 대응하는 입술 움직임이 고속으로 생성되는 것으로 가정한다.

- [0022] 음성과 입술 움직임이 서로 대응하지 않는 것이 분명하다: 입술 움직임이 중단되었을 때, 음성이 계속된다(오늘은 맑다...). 이것은, 공격자가 사용자의 ID 및 얼굴 검출을 통과하려고 시도한 경우, 공격자가 사용자(공격당한 사용자)의 이전 비디오 레코드를 사용하여 얼굴 검출을 공격할 수 있는 경우, 및 공격자가 사용자의 ID를 판독하여 음성 콘텐츠 ID의 인식을 공격하는 경우 등의 다수의 가능한 상황에서 발생할 수 있다. 이와 같이, 공격자가 인증될 수 있다. 그러나 그러한 공격에서, 입술 판독과 음성은 대개 일관성이 없으며, 판독자가 사용자가 아니라고 결정될 수 있다.
- [0023] 단계 102에서 도시된 바와 같이, 등록될 오디오 및 비디오 스트림에서의 입술 판독과 음성이 불일치하다는 것이 결정되면, 등록 실패가 사용자에게 통지될 수 있다. 대안적으로, 도 1에 도시된 바와 같이, 다시 오디오 및 비디오 스트림을 다시 수집하여 프로세싱 에러를 방지하도록 단계 101로 진행된다.
- [0024] 그렇지 않으면, 등록될 오디오 및 비디오 스트림에서의 입술 판독과 음성이 일치한다는 것이 결정되면, 단계 103을 수행한다: 수집된 오디오 및 비디오 스트림에서의 오디오 스트림에 대하여 음성 인식을 수행함으로써 취득된 음성 콘텐츠를 타겟 오브젝트의 오브젝트 식별자로 사용한다. 음성 인식은, 음성을 콘텐츠로 컨버팅하는 인식 프로세스인 사람의 음성 콘텐츠를 자동적으로 인식하기 위한 컴퓨터 기술을 사용한다. 예를 들어, 등록을 수행하기 위한 사용자에 의한 "123456"을 판독한 음성에 대하여 음성 인식이 수행된 후에, 오디오 스트림의 취득된 음성 콘텐츠는 "123456"이고, 인식을 통해 취득된 콘텐츠는 사용자의 식별자 즉 사용자의 ID로 사용될 수 있다.
- [0025] 오디오 스트림의 이전 음성 인식은, 오브젝트 식별자를 취득하기 위해 입술 판독과 음성이 일치하는 것으로 결정된 후에 수행되거나, 오브젝트 식별자를 취득하기 위해 입술 판독과 음성이 일치하는지 여부를 결정하는 프로세스에서 수행될 수 있다.
- [0026] 다른 양태에서, 등록될 오디오 및 비디오 스트림의 모델 생리 특징을 취득하기 위해, 등록될 오디오 및 비디오 스트림에 생리 인식이 수행된다. 본 실시예에서, 생리 특징은 얼굴 특징 및 성문 특징을 포함하지만, 예시적인 특징이 사용자를 고유하게 식별 할 수 있고, 상이한 사용자의 생리 특징을 구별하는데 사용될 수 있는 한, 2 개의 특징으로 제한되지 않는다. 본 양태에서, 단계 104에 도시된 바와 같이, 타겟 오브젝트의 음성 특징을 취득하기 위해, 등록될 오디오 및 비디오 스트림에서의 오디오 스트림에 대하여 성문 인식이 수행될 수 있다.
- [0027] 또 다른 양태에서, 타겟 오브젝트의 얼굴 특징을 취득하기 위해, 등록될 오디오 및 비디오 스트림에서의 비디오 스트림에 얼굴 검출이 수행된다.
- [0028] 등록 절차에서, 검출된 얼굴 특징은 모델 얼굴 특징으로 지칭될 수 있으며 후속 인증 과정에서 기준으로 사용된다. 마찬가지로, 검출된 성문 특징도 모델 성문 특징으로 지칭될 수 있으며, 모델 성문 특징 및 모델 얼굴 특징은 집합적으로 모델 생리 특징이라 지칭될 수 있다.
- [0029] 이러한 구현예에서, 모델 생리 특징 및 타겟 오브젝트의 오브젝트 식별자는 오브젝트 등록 정보로서 지칭될 수도 있다. 오브젝트 등록 정보의 데이터가 완성된 것으로 결정된 후에, 단계 106에서, 타겟 오브젝트의 오브젝트 식별자 및 대응하는 모델 생리 특징이 오브젝트 등록 정보로서 데이터베이스에 저장된다.
- [0030] 또한, 도 1에 도시된 3개의 양태의 실행 순서는 한정되지 않는다. 예를 들어, 등록될 오디오 및 비디오 스트림이 단계 101에서 취득된 후에, 3개의 양태는 병렬로 실행될 수 있다. 입술 판독과 음성이 불일치하면, 인식된 성문 특징과 인식된 얼굴 특징은 저장되지 않을 수 있다. 대안적으로, 입술 판독과 음성이 일치하는 것으로 결정된 후에, 성문 특징과 얼굴 특징에 대하여 검출 및 인식이 수행된다.
- [0031] 도 2는 도 1에서의 입술 판독과 음성 사이의 일치를 결정하는 절차를 예시한다. 상기 절차는 하기의 단계들을 포함할 수 있다:
- [0032] 단계 201: 등록될 오디오 및 비디오 스트림에서의 오디오 스트림에 기초하여 엔드포인트 검출을 수행한다. 현 단계에서, 오디오 스트림의 시작 시간과 종료 시간은 연속 오디오 스트림에서 검출될 수 있다.
- [0033] 단계 202: 오디오 스트림에 기초하여 연속 음성 특징을 추출하고, 이 특징은 MFCC 특징 및 LPCC 특징을 포함하지만 이것에 한정되지 않는다. 현 단계에서 추출되는 특징은 음성 인식을 위해 사용될 수 있다.
- [0034] 단계 203 : 오디오 스트림에서 음성 음절 및 대응하는 시간 포인트를 식별한다. 현 단계에서, 오디오 스트림에서의 각각의 음절은 단계 202에서 추출된 음성 특징에 기초하여 식별될 수 있고, 음절의 대응하는 출현 시간 포인트 및 대응하는 소멸 시간 포인트가 결정될 수 있다. 음성 인식 방법은 HMM(hidden Markov model), DNN(deep

neural network), 및 LSTM(Long Short Time Model) 등의 방법을 포함할 수 있지만, 이것에 한정되지 않는다.

- [0035] 단계 204: 등록될 오디오 및 비디오 스트림에서의 비디오 스트림에 기초하여 입술의 위치를 검출한다. 현 단계에서, 입술의 위치는 비디오 이미지로부터 검출될 수 있다.
- [0036] 단계 205: 검출된 입술 이미지의 품질을 결정한다. 예를 들어, 입술의 위치의 데피니션(definition) 및 노출과 같은 파라미터가 결정될 수 있다. 데피니션이 낮거나 노출 정도가 너무 높으면, 이미지가 검증되지 않은 것으로 결정되고, 이 경우 다시 등록될 오디오 및 비디오 스트림을 수집한다. 이미지의 품질이 충족되면, 단계 206으로 진행하여 입술 판독 인식을 수행한다.
- [0037] 단계 206: 입술의 연속 특징들을 추출한다. 현 단계에서, 특징은 연속 입술 이미지로부터 추출될 수 있고, 특징은 베어 픽셀(bare pixel), LBP, 가보(Gabor), SIFT, 또는 서프(Surf)와 같은 로컬 이미지 디스크립터(local image descriptor)를 포함하지만 이것에 한정되지 않는다.
- [0038] 단계 207: 비디오 스트림에서의 입술 판독 음절 및 대응하는 시간 포인트를 식별한다. 현 단계에서, 입술 판독 음절은 HMM(hidden Markov model) 및 Long Short Time Model과 같은 방법을 사용함으로써 식별될 수 있다. 입술 판독 인식 동안 모델을 사용함으로써 비디오 시간 시퀀스에서의 입술 판독 음절에 대응하는 시간 포인트가 또한 결정된다.
- [0039] 단계 208: 입술 판독 음절과 음성 음절이 모두 대응하는 시간 포인트와 일치하는지 여부를 결정한다. 예를 들어, 현 단계에서, 음성 음절의 시간 포인트 정보는 입술 판독 음절의 시간 포인트 정보와 비교될 수 있다. 비교 결과가 일치를 나타내면, 오디오 스트림이 실제 사람에 의해 생성된 것으로 간주되고; 이 경우, 단계 209로 진행된다. 비교 결과가 불일치를 나타내면 오디오 스트림이 공격 행위라고 의심되고; 이 경우 등록 절차로 돌아간다. 본 구현예에서, 입술 판독 음절과 음성 음절 및 대응하는 시간 포인트 사이의 일치를 검출하는 방법이 보다 상세하고, 이에 따라 실제 사람의 음성을 결정하는데 더 높은 정확도가 있다.
- [0040] 단계 209: 사용자의 ID, 즉 오브젝트 식별자를 취득하기 위해, 단계 202에서 추출된 음성 특성에 대하여 음성 인식을 수행한다. 음성 인식 방법은 HMM(hidden Markov model), DNN(deep neural network), 및 LSTM(Long Short Time Model) 등의 방법을 포함할 수 있지만, 이것에 한정되지 않는다.
- [0041] 또한, 도 2에 도시된 실시예에서, 입술 판독과 음성이 일치한다고 결정된 후에 오디오 스트림의 음성 인식이 단계 209에서 수행될 수 있다. 대안적으로, 오디오 스트림에서의 음절 시간 포인트가 단계 203에서 식별되면, 음성 특성에 기초하여 음성 인식을 수행함으로써 사용자의 ID가 취득된다. 이러한 경우에, 단계 208에서 입술 판독과 음성이 일치하는 것으로 결정된 후에, 인식을 통해 취득된 사용자의 ID가 오브젝트 식별자로서 직접 사용될 수 있다.
- [0042] 도 3은 도 1에서의 얼굴 특징 인식 절차를 예시한다. 상기 절차는 하기의 단계들을 포함할 수 있다:
- [0043] 단계 301: 등록될 오디오 및 비디오 스트림에서의 비디오 스트림에 기초하여 얼굴 이미지를 검출한다. 현 단계에서, 비디오 프레임 이미지는 오디오 및 비디오 스트림에서의 비디오 스트림으로부터 추출될 수 있고, 얼굴이 비디오 프레임 이미지에서 나타나는지 여부가 검출될 수 있다. 검출되면, 단계 302로 진행된다. 검출되지 않으면, 결정 절차로 다시 돌아간다.
- [0044] 단계 302: 얼굴 이미지의 품질을 검출한다. 현 단계에서, 단계 301에서 검출된 얼굴에 대하여 얼굴 특징 포인트 검출이 수행될 수 있고, 특징 포인트 검출이 결과에 기초하여 수평 방향 및 수직 방향 모두에서 얼굴의 각도가 결정될 수 있다. 두 각도가 특정 경사각을 초과하지 않으면 품질 요구사항이 충족된다. 그렇지 않으면, 품질 요구사항이 충족되지 않는다. 또한, 특정 임계값 내에 있어야 하는, 얼굴 영역의 데피니션, 노출 등이 결정된다. 얼굴 이미지의 품질이 양호하면, 얼굴 특징은 더 잘 인식될 수 있다.
- [0045] 단계 303 : 품질 요구사항을 만족시키는 얼굴 이미지에 대해, 고유벡터가 국부적인 이진 패턴(local binary pattern; LBP), Gabor, 콘볼루션 뉴럴 네트워크(convolutional neural network; CNN) 등을 포함하지만 이것에 한정되지 않는, 얼굴 이미지로부터 고유벡터를 추출한다.
- [0046] 단계 304 : 사용자의 고유의 얼굴 특징, 즉 모델 얼굴 특징을 형성하기 위해 단계 303에서 추출된 복수의 얼굴 고유벡터를 결합한다.
- [0047] 도 4는 도 1에서의 성문 특징 인식 절차를 예시한다. 상기 절차는 하기의 단계들을 포함할 수 있다.
- [0048] 단계 401: 등록될 오디오 및 비디오 스트림에서의 오디오 스트림을 취득한다.

- [0049] 본 실시예에서, 등록될 오디오 및 비디오 스트림에서의 오디오 스트림에 기초하여 성문 특징 인식이 수행될 수 있다.
- [0050] 단계 402: 오디오 스트림의 오디오 품질이 품질 기준을 만족하는 것을 결정한다.
- [0051] 현 단계에서, 오디오 품질이 결정될 수 있다. 수집된 오디오 스트림의 품질이 좋을수록 오디오에 대한 성문 인식을 더 효율적으로 수행할 수 있다. 따라서, 후속 성문 인식이 수행되기 전에, 오디오 스트림의 품질이 먼저 결정될 수 있다. 예를 들어, 오디오 스트림 내의 음성의 신호 강도 및 신호 대 잡음비와 같은 정보는 음성이 품질 기준을 충족시키는지 여부를 결정하기 위해 계산될 수 있다. 예를 들어, 품질 기준은, 신호 대 잡음비가 특정 범위 내에 있는 것이 되거나, 음성의 신호 강도가 강도 임계값보다 큰 것이 될 수 있다. 오디오 스트림의 품질이 충족되면, 단계 403으로 진행된다. 그렇지 않으면, 등록될 오디오 및 비디오 스트림을 다시 수집한다.
- [0052] 단계 403: 성문 고유벡터를 오디오 스트림으로부터 추출한다.
- [0053] 본 실시예에서, 등록될 복수의 오디오 및 비디오 스트림이 있을 수 있다. 예를 들어, 사용자는 사용자의 ID를 2회 판독할 수 있고, 대응하는 2개의 오디오 및 비디오 스트림이 수집된다. 현 단계에서, 각각의 오디오 및 비디오 스트림에서의 오디오 스트림의 성문 고유벡터가 추출될 수 있다. 고유벡터는 복수의 방법으로 추출될 수 있고, 세부사항은 간략함을 위해 생략된다. 예를 들면, 오디오 스트림의 음성 신호로부터 음성 특징 파라미터 MFCC(mel-frequency cepstral coefficient)가 추출될 수 있고, i-벡터(화자 인식 알고리즘) 또는 PLDA(probabilistic linear discriminant analysis, 즉 성문 인식을 위한 채널 보상 알고리즘)와 같은 방법을 사용함으로써 고유벡터가 계산된다.
- [0054] 단계 404: 복수의 오디오 스트림의 성문 고유벡터가 일치하는지 여부를 결정한다.
- [0055] 예를 들어, 사용자가 등록 중에 적어도 2번 사용자의 ID를 판독하는 경우, 대응하는 적어도 2개의 수집된 오디오 스트림이 있다. 복수의 오디오 스트림의 성문 특징들 사이의 차이가 너무 크지 않게 하기 위해, 복수의 오디오 스트림들 사이의 성문 일치 결정이 수행될 수 있다. 예를 들어, 복수의 오디오 스트림들 사이의 유사성 스코어(score of similarity)는 단계 403에서 각 오디오 스트림으로부터 추출된 성문 고유벡터에 기초하여 계산될 수 있다.
- [0056] 유사성 스코어가 특정 스코어 임계값 내에 있으면, 오디오 스트림이 유사성 요구사항을 충족시키는 것을 나타내고, 이 경우에, 단계 405로 진행된다. 그렇지 않으면, 사용자의 의해 입력된 복수의 오디오 사이에 큰 차이가 있다는 것을 나타내고, 등록을 수행하는 사용자에게 사용자는 ID를 다시 판독하도록, 즉 오디오 스트림을 다시 수집하도록 지시할 수 있다.
- [0057] 단계 405: 복수의 오디오 스트림의 성문 고유벡터에 기초하여 모델 성문 특징을 생성한다.
- [0058] 현 단계에서, 이전 단계에서 오디오 스트림으로부터 추출된 성문 고유벡터에 대해 가중 합계가 수행되어 모델 성문 특징을 취득할 수 있다.
- [0059] 이전의 등록 절차가 완료된 후에, 타겟 오브젝트의 오브젝트 등록 정보가 데이터베이스에 저장된다. 오브젝트 등록 정보는 오브젝트 식별자 및 대응하는 모델 생리 특징을 포함할 수 있다. 모델 생리 특징은 모델 성문 특징 및 모델 얼굴 특징을 포함할 수 있고, 오브젝트의 신원 인증 프로세싱은 오브젝트 등록 정보의 기반 아래에서 수행될 수 있다.
- [0060] [신원 인증]
- [0061] 도 5은 예시적 신원 인증 절차를 예시한다. 본 절차에서, 인증에 사용되는 생리 특징은 실시예로서 얼굴 특징과 성문 특징의 조합을 사용하여 설명된다. 또한, 생리 특징은 인증되는 타겟 오브젝트가 비디오 대신에 살아있는 오브젝트인 것으로 결정된 후에 비교될 수 있다. 도 5에 도시된 바와 같이, 인증 절차는 하기 프로세싱을 포함한다:
- [0062] 단계 501: 수집된 오디오 비디오 스트림을 취득하고, 오디오 비디오 스트림은 인증될 타겟 오브젝트에 의해 생성된다.
- [0063] 예를 들어, 사용자의 신원이 애플리케이션에 의해 인증된 후에만, 비교적 높은 보안 요구사항을 가진 애플리케이션에 사용자가 로그인할 수 있다. 현 단계에서, 사용자는 예를 들어 사용자의 스마트폰과 같은 지능형 디바이스 상에서 애플리케이션을 시작할 수 있고, 사용자는 스마트폰의 카메라 및 마이크를 사용하여 인증될 오디오 및 비디오 스트림을 수집할 수 있다. 사용자가 사용자의 애플리케이션 ID를 판독할 때 오디오 및 비디오 스트림

이 생성될 수 있다.

- [0064] 단계 502: 오디오 및 비디오 스트림에서의 입술 판독과 음성이 일치하는지 여부를 결정한다.
- [0065] 본 실시예에서, 오디오 및 비디오 스트림에서의 입술 판독과 음성이 일치하는지 여부가 먼저 결정될 수 있다. 특유의 일치 결정 절차에 대하여 도 2를 참조하고, 세부사항은 간략함을 위해 생략된다.
- [0066] 입술 판독과 음성이 일치하면, 인증되는 타겟 오브젝트가 비디오 등이 아닌 살아있는 오브젝트라는 것을 나타낸다. 이 경우에, 단계 503으로 진행된다. 그렇지 않으면, 다시 수집을 수행하기 위해 단계 501로 돌아간다.
- [0067] 단계 503: 오디오 스트림의 음성 콘텐츠를 취득하기 위해 오디오 및 비디오 스트림에서의 오디오 스트림에 음성 인식을 수행한다. 예를 들어, 인식을 통해 취득된 음성 콘텐츠는 사용자의 ID "123456"이 될 수 있다.
- [0068] 단계 504: 타겟 오브젝트의 오브젝트 식별자로서 음성 콘텐츠를 사용하고, 미리 저장된 오브젝트 등록 정보가 오브젝트 식별자를 포함하는지 여부를 결정한다.
- [0069] 예를 들어, 미리 저장된 오브젝트 등록 정보가 오브젝트 식별자를 포함하면, 오브젝트 식별자에 대응하는 예를 들어 모델 얼굴 특징 및 모델 성문 특징과 같은 모델 생리 특징은 오브젝트 등록 정보에서 취득될 수 있다. 또한, 생리 특징을 모델 생리 특징과 비교하기 위해, 타겟 오브젝트의 생리 특징을 취득하도록 인증될 오디오 및 비디오 스트림에 대하여 생리 인식이 수행된다. 미리 저장된 오브젝트 등록 정보가 오브젝트 식별자를 포함하지 않으면, 사용자가 등록을 수행하지 않았다고 사용자에게 통지될 수 있다.
- [0070] 단계 505: 타겟 오브젝트의 성문 특징을 취득하기 위해 오디오 및 비디오 스트림에 대하여 성문 인식을 수행한다. 현 단계에서의 성문 특징의 추출에 대하여, 도 4를 참조한다.
- [0071] 단계 506: 타겟 오브젝트의 얼굴 특징을 취득하기 위해 오디오 및 비디오 스트림에 대하여 얼굴 인식을 수행한다.
- [0072] 이어서, 타겟 오브젝트의 생리 특징이 모델 생리 특징과 비교되어 비교 결과가 취득될 수 있고, 비교 결과가 인증 조건을 충족시키면 타겟 오브젝트가 인증되었다고 결정된다. 예를 들어, 단계 507 내지 단계 509가 포함된다.
- [0073] 단계 507: 성문 비교 스코어를 취득하기 위해 타겟 오브젝트의 성문 특징을 모델 성문 특징과 비교한다.
- [0074] 단계 508: 얼굴 비교 스코어를 취득하기 위해 타겟 오브젝트의 얼굴 특징을 모델 얼굴 특징과 비교한다.
- [0075] 단계 509: 성문 비교 스코어 및 얼굴 정렬 스코어가 인증 조건을 충족시키는지 여부를 결정한다.
- [0076] 예컨대, 성문 비교 스코어와 얼굴 비교 스코어가, 성문 비교 스코어가 성문 스코어 임계값보다 크고, 얼굴 비교 스코어가 얼굴 스코어 임계값보다 큰 조건; 또는 성문 비교 스코어와 얼굴 비교 스코어의 곱(product)이 대응하는 곱 임계값보다 큰 조건; 또는 성문 비교 스코어와 얼굴 비교 스코어의 가중 합계가 대응하는 가중 임계값보다 큰 조건 중 적어도 하나를 충족시키면, 타겟 오브젝트가 인증된 것으로 결정된다.
- [0077] 성문 비교 스코어와 얼굴 비교 스코어가 현 단계에서 인증 조건을 충족시키는 것으로 결정되면, 타겟 오브젝트가 인증된 것으로 결정된다. 그렇지 않으면, 타겟 오브젝트는 인증에 실패한 것으로 결정된다.
- [0078] 또한, 이 신원 인증의 실시예에서는 이전의 신원 등록 절차와 마찬가지로, 입술 판독과 음성이 일치하는 것으로 결정된 후에 사용자의 ID를 취득하기 위해 오디오 스트림에 대하여 음성 인식이 수행되거나, 오디오 스트림에서의 음절 시간 포인트가 식별될 때 사용자의 ID가 취득될 수 있다. 이전 실시예에서, 입술 판독과 음성이 일치하는 것으로 결정된 후에 사용자의 ID가 가 식별된다.
- [0079] 본 출원의 이 구현예에서의 신원 인증 방법에 따르면, 오디오 및 비디오 스트림은 사용자의 인증 중에 한번만 생성될 필요가 있다. 예를 들어, 사용자는 사용자의 ID를 한번만 판독할 필요가 있다. 상기 방법에서, 오디오 및 비디오 스트림에 대하여 음성 인식을 수행함으로써 사용자의 ID가 취득될 수 있고, 얼굴 특징 및 성문 특징도 동일한 오디오 및 비디오 스트림을 사용함으로써 검증될 수 있다. 이것은, 사용자의 작업을 단순화하고, 인증 효율을 높이며, 일대일 인증 모델을 유지 관리하고, 인식 정확도를 보장한다. 한편, 인식된 생리 특징은 데이터베이스 내의 오브젝트 식별자에 대응하는 특징과만 비교되고, 이에 따라 인식 정확도가 보장된다. 또한, 이 방법에서는 입술 판독과 음성의 일치가 결정되어, 타겟 오브젝트가 공격자의 가짜 비디오 레코드가 아닌 살아있는 오브젝트임을 보장함으로써, 인증 보안 및 안정성을 향상시킨다. 상기 방법에서, 인식을 통해 취득된 생리 특징 및 사용자의 ID는 동일한 오디오 및 비디오 스트림에 기초하여 취득된다. 어느 정도까지는, 공격자의

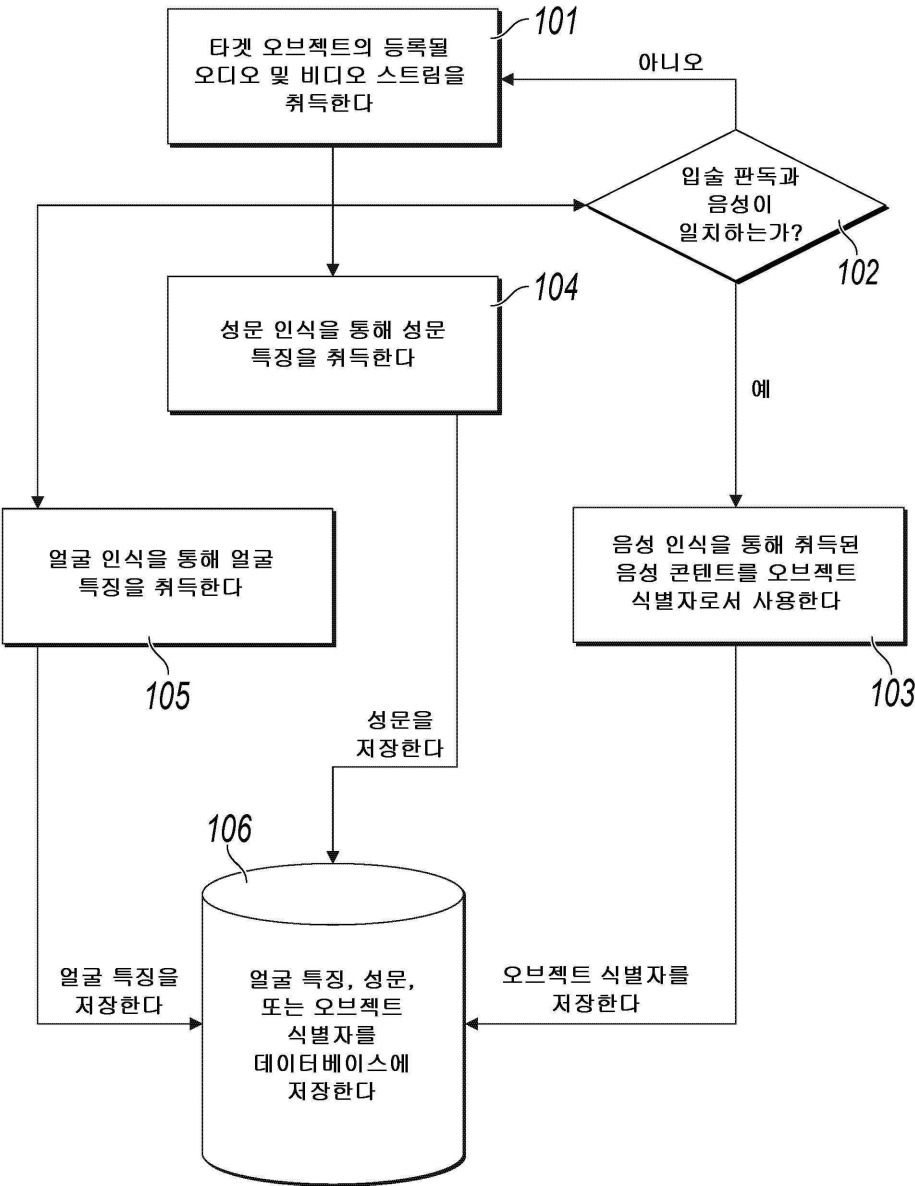
가짜 오디오 및 비디오 스트림이 인식될 수 있다.

- [0080] 이전의 신원 인증 방법을 구현하기 위해, 본 출원의 구현에는 또한 신원 인증 장치를 제공한다. 도 6에 도시된 바와 같이, 상기 장치는 정보 취득 모듈(61), 식별자 결정 모듈(62), 정보 관리 모듈(63), 특징 인식 모듈(64), 및 인증 프로세싱 모듈(65)을 포함할 수 있다.
- [0081] 정보 취득 모듈(61)은, 수집된 오디오 및 비디오 스트림을 취득하도록 구성되고, 오디오 및 비디오 스트림은 인증될 타겟 오브젝트에 의해 생성된다.
- [0082] 식별자 결정 모듈(62)은, 오디오 비디오 스트림 내의 입술 판독과 음성이 일치하는지 여부를 결정하고, 입술 판독과 음성이 일치하면, 오디오 및 비디오 스트림에서의 오디오 스트림에 대하여 음성 인식을 수행함으로써 취득된 음성 콘텐츠를 타겟 오브젝트의 오브젝트 식별자로서 사용하도록 구성된다.
- [0083] 정보 관리 모듈(63)은, 미리 저장된 오브젝트 등록 정보가 오브젝트 식별자를 포함하면, 오브젝트 등록 정보로부터 오브젝트 식별자에 대응하는 모델 생리 특징을 취득하도록 구성된다.
- [0084] 특징 인식 모듈(64)은, 타겟 오브젝트의 생리 특징을 취득하기 위해 오디오 및 비디오 스트림에 대하여 생리 인식을 수행하도록 구성된다.
- [0085] 인증 프로세싱 모듈(65)은, 타겟 오브젝트의 생리 특징을 모델 생리 특징과 비교하여 비교 결과를 취득하고, 비교 결과가 인증 조건을 충족시키면, 타겟 오브젝트가 인증된 것으로 결정하도록 구성된다.
- [0086] 도 7을 참조하면, 실시예에서, 특징 인식 모듈(64)은 성문 인식 서브모듈(641) 및 얼굴 인식 서브모듈(642)을 포함할 수 있다.
- [0087] 성문 인식 서브모듈(641)은 타겟 오브젝트의 성문 특징을 취득하기 위해 오디오 및 비디오 스트림에 대하여 성문 인식을 수행하도록 구성된다.
- [0088] 얼굴 인식 서브모듈(642)은 타겟 오브젝트의 얼굴 특징을 취득하기 위해 오디오 및 비디오 스트림에 대하여 얼굴 인식을 수행하도록 구성된다.
- [0089] 인증 프로세싱 모듈(65)은, 타겟 오브젝트의 성문 특징을 모델 성문 특징과 비교하여 성문 비교 스코어를 취득하고, 타겟 오브젝트의 얼굴 특징을 모델 얼굴 특징과 비교하여 얼굴 비교 스코어를 취득하며; 성문 비교 스코어와 얼굴 비교 스코어가 인증 조건을 충족시키면, 타겟 오브젝트가 인증되었다고 결정하도록 구성된다.
- [0090] 실시예에서, 성문 비교 스코어와 얼굴 비교 스코어가, 성문 비교 스코어가 성문 스코어 임계값보다 크고, 얼굴 비교 스코어가 얼굴 스코어 임계값보다 큰 조건; 또는 성문 비교 스코어와 얼굴 비교 스코어의 곱(product)이 대응하는 곱 임계값보다 큰 조건; 또는 성문 비교 스코어와 얼굴 비교 스코어의 가중 합계가 대응하는 가중 임계값보다 큰 조건 중 적어도 하나를 충족시키면, 타겟 오브젝트가 인증된 것으로 결정된다.
- [0091] 실시예에서, 도 7에 도시된 바와 같이, 식별자 결정 모듈(62)은, 오디오 및 비디오 스트림에서의 오디오 스트림 내의 음성 음절과 대응하는 시간 포인트를 식별하고, 오디오 및 비디오 스트림에서의 비디오 스트림 내의 입술 판독 음절과 대응하는 시간 포인트를 식별하도록 구성된 음성 인식 서브모듈(621); 및 음성 음절과 입술 판독 음절이 대응하는 시간 포인트와 일치하면 입술 판독과 음성이 일치한다고 결정하도록 구성된 일치 결정 서브모듈(622)을 포함할 수 있다.
- [0092] 실시예에서, 정보 취득 모듈(61)은 또한, 타겟 오브젝트의 등록될 오디오 및 비디오 스트림을 취득하도록 구성된다.
- [0093] 식별자 결정 모듈(62)은 또한, 등록될 오디오 및 비디오 스트림에서의 입술 판독과 음성이 일치할 때, 오디오 및 비디오 스트림에서의 오디오 스트림에 대하여 음성 인식을 수행함으로써 취득된 음성 콘텐츠를 타겟 오브젝트의 오브젝트 식별자로서 사용하도록 구성된다.
- [0094] 특징 인식 모듈(64)은 또한, 등록될 오디오 및 비디오 스트림의 모델 생리 특징을 취득하기 위해 등록될 오디오 및 비디오 스트림에 대하여 생리 인식을 수행하도록 구성된다.
- [0095] 정보 관리 모듈(63)은 또한, 타겟 오브젝트의 오브젝트 식별자 및 오브젝트 등록 정보 내의 대응하는 모델 생리 특징을 대응적으로 저장하도록 구성된다.
- [0096] 전술한 설명은 본 출원의 단지 예시적인 구현예일 뿐이며, 본 출원을 제한하려는 것은 아니다. 본 출원의 사상

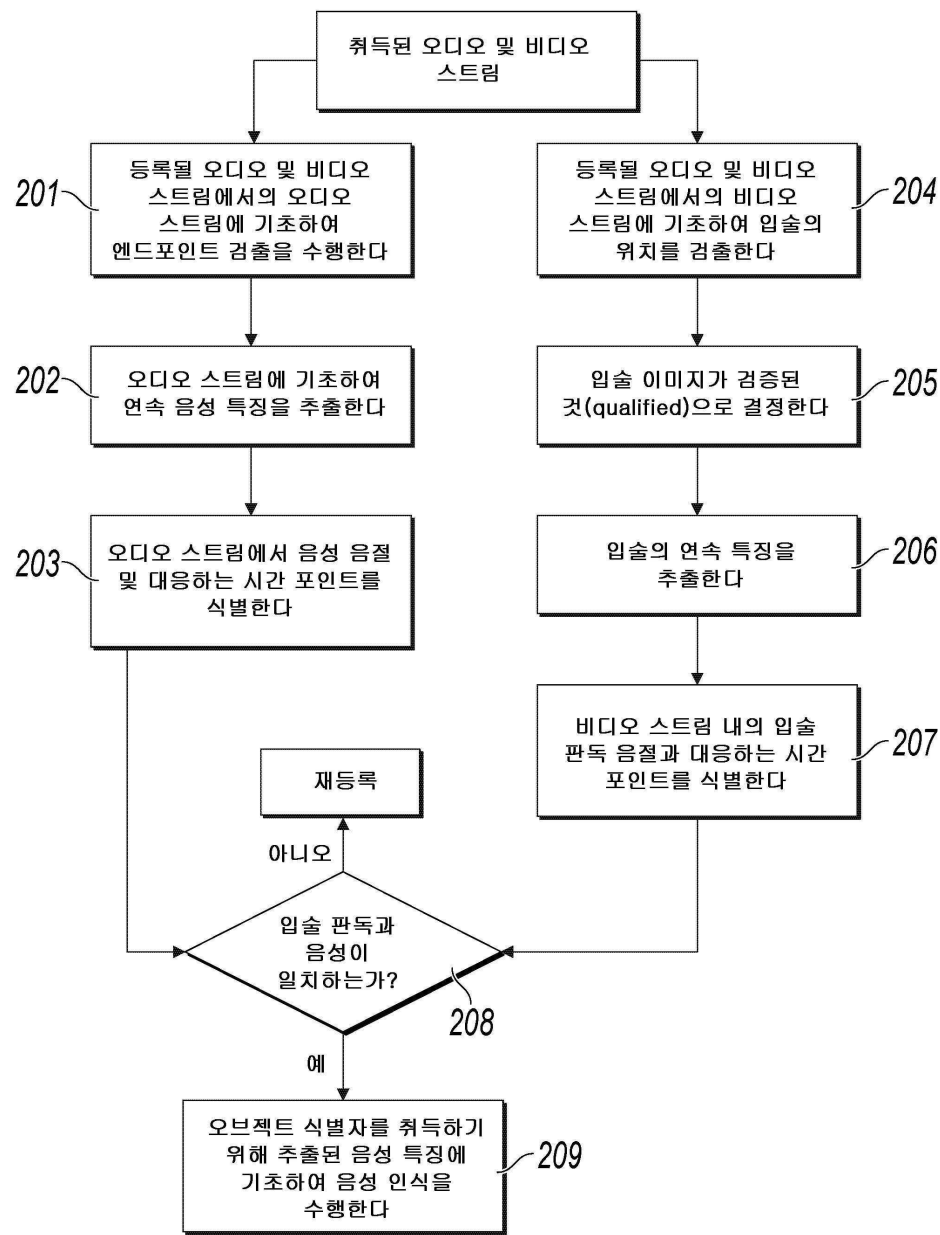
및 원리 내에서 이루어진 임의의 수정, 등가의 치환, 개선 등은 본 출원의 보호 범위 내에 있다.

도면

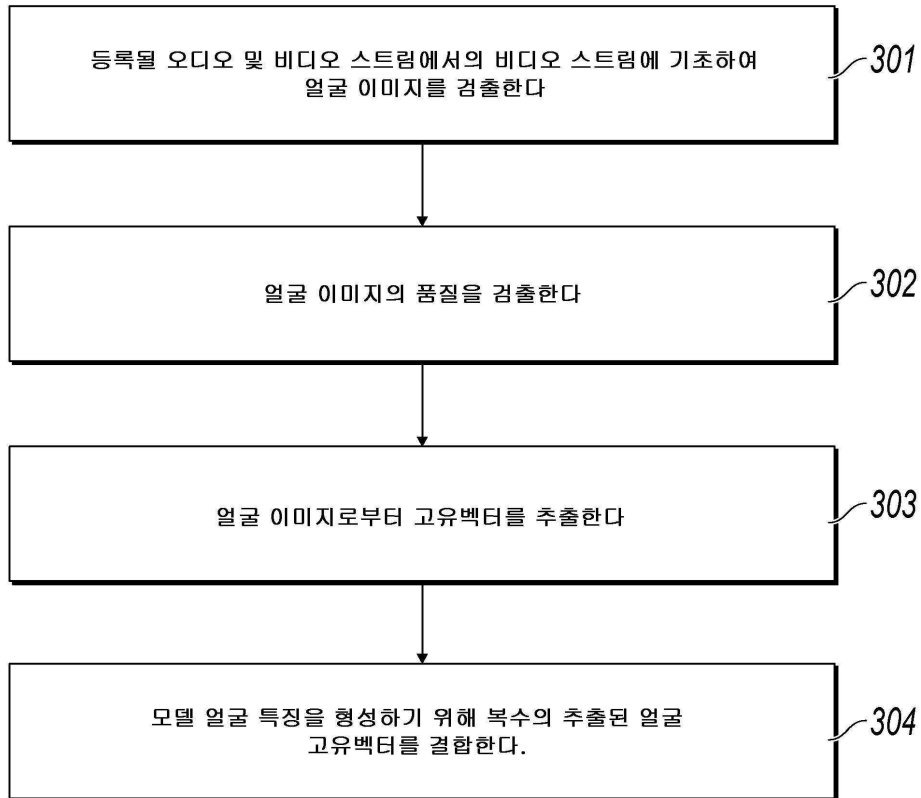
도면1



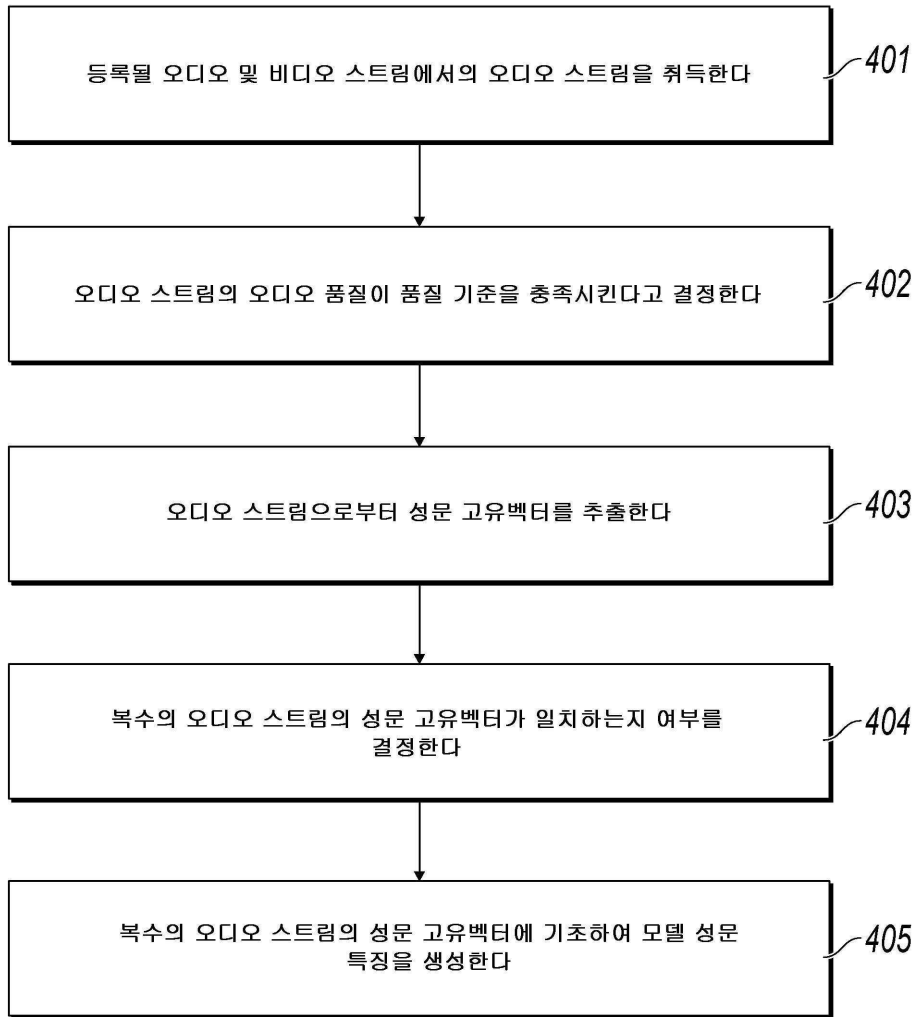
도면2



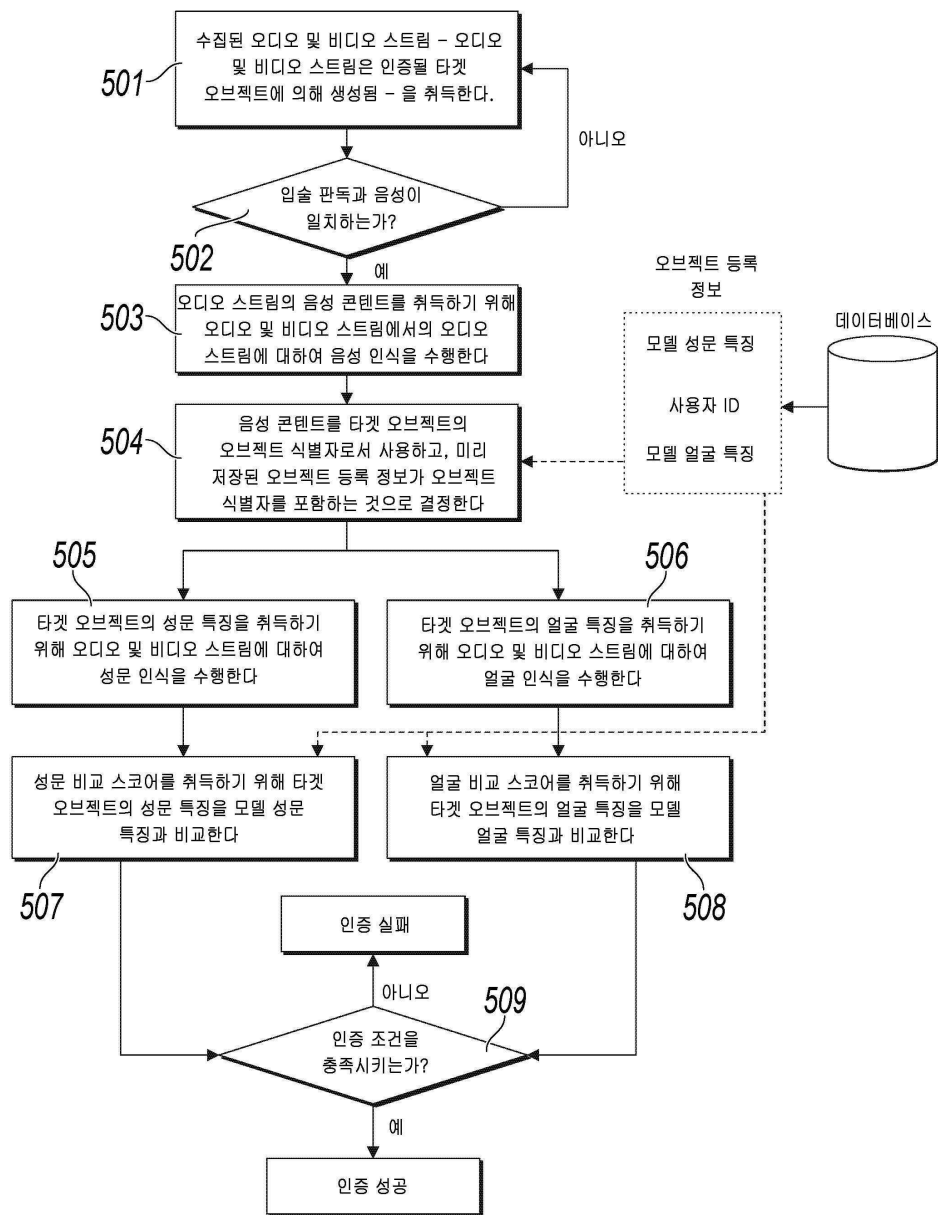
도면3



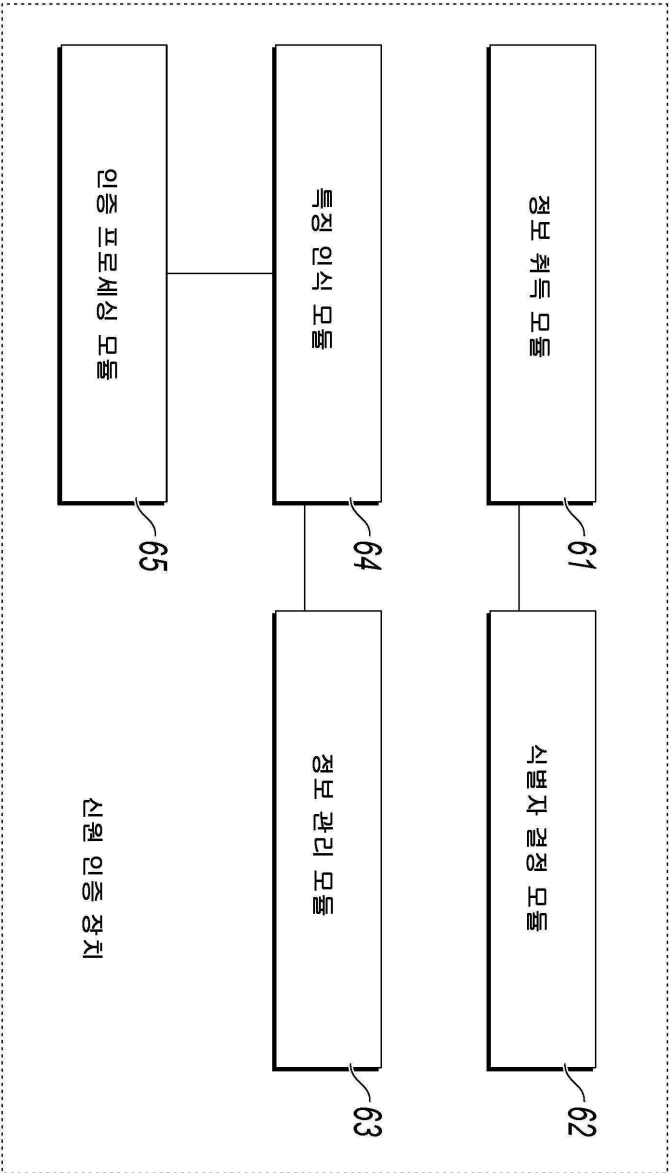
도면4



도면5



도면6



도면7

