



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2013년06월20일
(11) 등록번호 10-1276204
(24) 등록일자 2013년06월12일

(51) 국제특허분류(Int. Cl.)
G06T 7/00 (2006.01) G10L 17/04 (2013.01)
(21) 출원번호 10-2010-0044142
(22) 출원일자 2010년05월11일
심사청구일자 2010년05월11일
(65) 공개번호 10-2011-0124644
(43) 공개일자 2011년11월17일
(56) 선행기술조사문헌
KR1020070075644 A
KR1020080049394 A
KR1020080100097 A
KR1020090107666 A

(73) 특허권자
한국전자통신연구원
대전광역시 유성구 가정로 218 (가정동)
(72) 발명자
김혜진
대전광역시 유성구 가정로 65, 두레APT 103동 70
5호 (신성동)
김도형
대전광역시 유성구 가정로 65, 110동 602호 (신성
동, 대림두레아파트)
(뒷면에 계속)
(74) 대리인
특허법인지명

전체 청구항 수 : 총 15 항

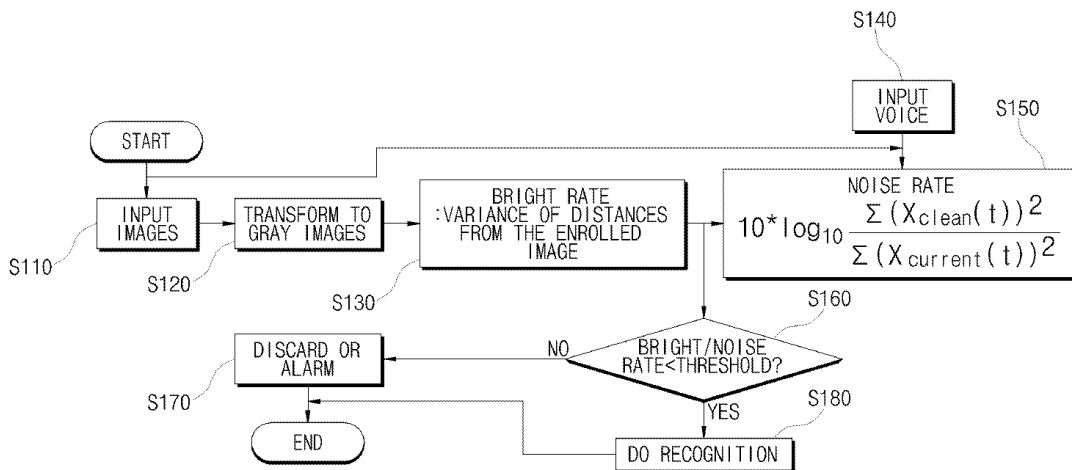
심사관 : 진민속

(54) 발명의 명칭 멀티모달 융합을 위한 환경변수 측정방법

(57) 요약

멀티모달 융합을 위한 환경변수 측정방법이 제공된다. 환경변수 측정방법은 적어도 하나 이상의 등록 모달리티를 준비하는 단계, 적어도 하나 이상의 입력 모달리티를 받는 단계, 하나의 등록 모달리티로서의 등록 영상의 조도를 기준으로 하나의 입력 모달리티로서의 입력 영상에 대한 영상관련환경변수를 계산하는 단계, 및 영상관련환경변수를 기설정된 기준값과 비교하고 그 비교 결과에 따라 입력 영상을 폐기하거나 또는 인식 데이터로서 출력하는 단계를 포함한다.

대표도



(72) 발명자

지수영

대전광역시 유성구 궁동로 5, 102동 305호 395-3
(궁동, 다솔아파트)

이재연

대전광역시 유성구 어은로 57, 131동 1501호 (어은
동, 한빛아파트)

이 발명을 지원한 국가연구개발사업

과제고유번호 2008-F-037-01

부처명 지식경제부

연구사업명 IT원천기술개발

연구과제명 u-로봇 HRI 솔루션 및 핵심 소자 기술 개발

주관기관 한국전자통신연구원

연구기간 2008년 03월 01일 ~ 2012년 02월 29일

특허청구의 범위

청구항 1

적어도 하나의 등록 모달리티를 준비하는 제1 단계;

적어도 하나의 입력 모달리티를 받는 제2 단계;

상기 적어도 하나의 등록 모달리티 중 등록 영상의 조도를 기준으로 상기 적어도 하나의 입력 모달리티 중 입력 영상의 영상관련환경변수를 계산하는 제3 단계; 및

상기 영상관련환경변수를 기설정된 기준값과 비교하고, 비교 결과에 따라 상기 입력 영상을 폐기하거나 또는 인식 데이터로서 출력하는 제4 단계

를 포함하는 멀티모달 융합을 위한 환경변수 측정방법.

청구항 2

제1항에 있어서,

상기 입력 영상을 그레이 영상으로 변환하는 단계

를 더 포함하는 멀티모달 융합을 위한 환경변수 측정방법.

청구항 3

제2항에 있어서,

상기 제3 단계는 상기 등록 영상과 상기 입력 영상 간의 디스턴스 놈(distance norm)을 구하는 단계를 포함하는 멀티모달 융합을 위한 환경변수 측정방법.

청구항 4

제3항에 있어서,

상기 디스턴스 놈(distance norm)은 absolute distance(1-norm distance), Euclidean distance(2-norm distance), Minkowski distance(p-norm distance), Chebyshev distance, Mahalanobis distance, hamming distance, Lee distance, Levenshtein distance, 또는 이들의 조합을 포함하는 멀티모달 융합을 위한 환경변수 측정방법.

청구항 5

제1항에 있어서,

상기 등록 모달리티는 사용자 인식을 위한 상기 입력 영상의 비교 기준이 되는 상기 등록 영상과, 또 다른 입력 모달리티로서의 입력 음성의 비교 기준이 되는 등록 음성을 포함하는 멀티모달 융합을 위한 환경변수 측정방법.

청구항 6

제5항에 있어서,

상기 입력 음성에 대하여 수학식 2에 의해 음성관련환경변수(NoiseRate)를 구하는 단계를 더 포함하는 멀티모달 융합을 위한 환경변수 측정방법:

[수학식 2]

$$NoiseRate = 10 * \log \frac{\sum (x_{clean}(t))^2}{(x_{current}(t))^2}$$

(여기서, Xclean(t)는 사용자를 등록하는 환경에서의 상기 등록 음성을 나타내고, Xcurrent(t)는 임의의 환경에

서의 상기 입력 음성을 나타냄).

청구항 7

사용자 인식을 위한 등록 음성을 준비하는 제1 단계;
 상기 사용자 인식을 위한 입력 음성을 받는 제2 단계;
 상기 등록 음성을 기준으로 상기 입력 음성에 대한 음성관련환경변수를 추출하는 제3 단계; 및
 추출된 상기 음성관련환경변수를 기설정된 기준값과 비교하고, 비교 결과에 따라 상기 입력 음성을 폐기하거나 인식 데이터로서 출력하는 제4 단계
 를 포함하는 멀티모달 융합을 위한 환경변수 측정방법.

청구항 8

제7항에 있어서,
 상기 음성관련환경변수(NoiseRate)를 수학적 식 2에 따라 구하는 단계
 를 더 포함하는 멀티모달 융합을 위한 환경변수 측정방법:

[수학적 식 2]

$$NoiseRate = 10 * \log \frac{\sum (x_{clean}(t))^2}{(x_{current}(t))^2}$$

(여기서, Xclean(t)는 사용자를 등록하는 환경에서의 상기 등록 음성을 나타내고, Xcurrent(t)는 임의의 환경에서 상기 입력 음성을 나타냄).

청구항 9

제7항에 있어서,
 상기 제1 단계는 20dB 이상의 SNR 환경에서 상기 등록 음성을 준비하는 단계
 를 포함하는 멀티모달 융합을 위한 환경변수 측정방법.

청구항 10

사용자 인식을 위한 등록 영상 및 등록 음성을 준비하는 제1 단계;
 상기 사용자 인식을 위한 입력 영상 및 입력 음성을 각각 받는 제2 단계;
 상기 등록 영상을 기준으로 상기 입력 영상에 대한 영상관련환경변수를 추출하는 제3 단계;
 상기 등록 음성을 기준으로 상기 입력 음성에 대한 음성관련환경변수를 추출하는 제4 단계; 및
 추출된 상기 영상관련환경변수와 상기 음성관련환경변수를 기설정된 기준값과 각각 비교하고, 비교 결과에 따라 상기 입력 영상만, 상기 입력 음성만, 또는 상기 입력 영상과 상기 입력 음성 모두를 폐기하거나 인식 데이터로서 출력하는 제5 단계를 포함하는 멀티모달 융합을 위한 환경변수 측정방법.

청구항 11

제10항에 있어서,
 상기 입력 영상을 그레이 영상으로 변환하는 단계
 를 더 포함하는 멀티모달 융합을 위한 환경변수 측정방법.

청구항 12

제10항에 있어서,

상기 제3 단계는 수학적 식 1에 의해 상기 등록 영상과 상기 입력 영상 간의 디스턴스 놈(distance norm)을 연산하는 단계

를 포함하는 멀티모달 융합을 위한 환경변수 측정방법:

[수학적 식 1]

$$BrightRate = variance(distNorm(I_{enroll}, I_{test}))$$

(여기서, I_{enroll} 은 등록 영상을 나타내고, I_{test} 는 테스트 영상 또는 입력 영상을 나타내며, 연산된 상기 디스턴스 놈 값의 변화량(variance)은 상기 입력 영상에 대한 환경 변수인 BrightRate을 나타냄).

청구항 13

제12항에 있어서,

상기 디스턴스 놈(distance norm)은 absolute distance(1-norm distance), Euclidean distance(2-norm distance), Minkowski distance(p-norm distance), Chebyshev distance, Mahalanobis distance, hamming distance, Lee distance, Levenshtein distance, 또는 이들의 조합을 포함하는 멀티모달 융합을 위한 환경변수 측정방법.

청구항 14

제10항에 있어서,

상기 제4 단계는 상기 음성관련환경변수(NoiseRate)를 수학적 식 2에 따라 구하는 단계를 더 포함하는 멀티모달 융합을 위한 환경변수 측정방법:

[수학적 식 2]

$$NoiseRate = 10 * \log \frac{\sum (x_{clean}(t))^2}{(x_{current}(t))^2}$$

(여기서, $x_{clean}(t)$ 는 사용자를 등록하는 환경에서의 상기 등록 음성을 나타내고, $x_{current}(t)$ 는 임의의 환경에서의 상기 입력 음성을 나타냄).

청구항 15

제14항에 있어서,

상기 제1 단계는 20dB 이상의 SNR 환경에서 상기 등록 음성을 준비하는 단계

를 포함하는 멀티모달 융합을 위한 환경변수 측정방법.

명세서

기술분야

[0001] 본 발명은 멀티모달 융합을 위한 환경변수 측정방법에 관한 것이다.

[0002] 본 발명은 지식경제부 IT원천기술개발 사업의 일환으로 수행한 연구로부터 도출된 것이다[과제관리번호: 2008-F-037-01, 과제명: u-로봇 HRI 솔루션 및 핵심 소자 기술 개발].

배경기술

[0003] 종래의 멀티모달 융합 사용자 인식(multi-modal fusion user recognition) 방법은 복수의 멀티모달 정보를 인식물로 융합하거나 특징값(feature)으로 융합하는 방법들이 주를 이루고 있다. 융합의 목적이 여러 데이터를 합하여 더 나은 성능을 얻는 것이라고 할 때, 각각의 인체의 감각적 양상 즉 모달리티(modality) 데이터의 특성상, 그 인식률이 낮아지는 환경은 각기 서로 다르다. 예를 들면, 얼굴 인식물의 경우 역광과 같은 조건에서

인식률이 낮아지고, 화자 인식률의 경우 SNR(signal-to-noise ratio)이 높은 조건에서 인식률이 낮아진다.

[0004] 이와 같이, 사용자 인식에 있어서 인식률이 낮아지는 환경은 잘 알려져 있다. 하지만, 이러한 환경 변수를 사용자 인식 시스템에 반영하여 사용자 인식 성능을 높이지는 못하고 있다. 그것은 사용자 인식에 있어서 시시각각 달라지는 환경을 인식률에 영향을 주는 변수로서 측정하기가 어렵기 때문이다.

발명의 내용

해결하려는 과제

[0005] 본 발명의 목적은 실제 환경에서의 입력 영상, 입력 음성, 또는 이들 모두의 신뢰도를 실시간으로 측정할 수 있는 멀티모달 융합을 위한 환경변수 측정방법을 제공하는 데 있다.

과제의 해결 수단

[0006] 상기 기술적 과제를 해결하기 위하여 본 발명의 일 측면에 따르면, 적어도 하나 이상의 등록 모달리티를 준비하는 제1 단계; 적어도 하나 이상의 입력 모달리티를 받는 제2 단계; 적어도 하나 이상의 등록 모달리티 중 등록 영상의 조도를 기준으로 적어도 하나 이상의 입력 모달리티 중 입력 영상에 대한 영상관련환경변수를 계산하는 제3 단계; 및 영상관련환경변수를 기설정된 기준값과 비교하고, 비교 결과에 따라 입력 영상을 폐기하거나 또는 인식 데이터로서 출력하는 제4 단계를 포함하는 멀티모달 융합을 위한 환경변수 측정방법이 제공된다.

[0007] 본 발명의 또 다른 측면에 따르면, 사용자 인식을 위한 등록 음성을 준비하는 제1 단계; 사용자 인식을 위한 입력 음성을 받는 제2 단계; 등록 음성을 기준으로 입력 음성에 대한 음성관련환경변수를 추출하는 제3 단계; 및 추출된 음성관련환경변수를 기설정된 기준값과 비교하고, 비교 결과에 따라 입력 음성을 폐기하거나 인식 데이터로서 출력하는 제4 단계를 포함하는 멀티모달 융합을 위한 환경변수 측정방법이 제공된다.

[0008] 본 발명의 또 다른 측면에 따르면, 사용자 인식을 위한 등록 영상 및 등록 음성을 준비하는 제1 단계; 사용자 인식을 위한 입력 영상 및 입력 음성을 각각 받는 제2 단계; 등록 영상을 기준으로 입력 영상에 대한 영상관련 환경변수를 추출하는 제3 단계; 등록 음성을 기준으로 입력 음성에 대한 음성관련환경변수를 추출하는 제4 단계; 및 추출된 영상관련환경변수와 음성관련환경변수를 기설정된 기준값과 각각 비교하고, 비교 결과에 따라 입력 영상만, 입력 음성만, 또는 입력 영상과 입력 음성 모두를 폐기하거나 인식 데이터로서 출력하는 제5 단계를 포함하는 멀티모달 융합을 위한 환경변수 측정방법이 제공된다.

발명의 효과

[0009] 본 발명의 실시예에 따르면, 실제 환경에서 음성, 영상, 또는 이들 모두의 퀄리티(quality)를 실시간으로 측정할 수 있는 멀티모달 융합을 위한 환경변수 측정방법을 제공할 수 있다. 즉, 환경을 직접 측정하는 기존의 방식과 달리, 등록된 사용자 인식 정보를 토대로 입력된 인식 데이터의 사용자 환경을 실시간으로 측정하기 때문에 멀티모달 융합에서 사용자 인식에 대한 가중치나 파라미터로서 사용할 수 있다. 다시 말하면, 사용자 인식에 있어서 입력 데이터의 퀄리티에 대한 신뢰도를 제공하는 방안이 될 수 있다. 게다가, 극단적으로 좋지 않은 입력 데이터의 경우, 입력된 인식 데이터의 폐기나 새로운 인식 데이터의 입력을 간단히 판단할 수 있으므로, 상호작용이 가능한 사용자 인식 시스템에서 시스템의 속도 향상이나 불필요한 연산 방지 등에 유용하게 사용될 수 있다.

도면의 간단한 설명

[0010] 도 1은 본 발명의 일 실시예에 따른 멀티모달 융합을 위한 환경변수 측정방법에 대한 순서도이다.

도 2는 도 1의 멀티모달 융합을 위한 환경변수 측정방법에 채용가능한 등록 얼굴 영상의 일례를 나타내는 도면이다.

도 3은 도 1의 멀티모달 융합을 위한 환경변수 측정방법에 있어서 다양한 입력 영상에 대한 얼굴 인식 과정을 설명하기 위한 도면이다.

도 4는 도 3의 다양한 입력 영상에 대한 밝기를 설명하기 위한 도면이다.

도 5는 도 1의 멀티모달 융합을 위한 환경변수 측정방법에 있어서 조명 거리에 따른 BrightRate을 설명하기 위한 그래프이다.

도 6은 도 1의 멀티모달 융합을 위한 환경변수 측정방법에 있어서, BrightRate에 따른 인식 에러율을 나타내는 그래프이다.

발명을 실시하기 위한 구체적인 내용

- [0011] 이하, 첨부한 도면들 및 후술되는 내용을 참조하여 본 발명의 바람직한 실시예들을 상세히 설명한다. 그러나, 본 발명은 여기서 설명되는 실시예들에 한정되지 않고 다른 형태로 구체화될 수도 있다. 오히려, 여기서 소개되는 실시예들은 개시된 내용이 철저하고 완전해질 수 있도록 그리고 당업자에게 본 발명의 사상이 충분히 전달될 수 있도록 하기 위해 제공되는 것이다. 명세서 전체에 걸쳐서 동일한 참조번호들은 동일한 구성요소들을 나타낸다. 한편, 본 명세서에서 사용된 용어는 실시예들을 설명하기 위한 것이며 본 발명을 제한하고자 하는 것은 아니다. 본 명세서에서, 단수형은 문구에서 특별히 언급되지 않는 한 복수형도 포함된다. 명세서에서 사용되는 용어 "포함한다(comprise)" 및/또는 "포함하는(comprising)"은 언급된 구성요소, 단계, 동작 및/또는 소자가 하나 이상의 다른 구성요소, 단계, 동작 및/또는 소자의 존재 또는 추가를 배제하지 않는다.
- [0012] 도 1은 본 발명의 일 실시예에 따른 멀티모달 융합을 위한 환경변수 측정방법에 대한 순서도이다.
- [0013] 이하의 설명에서, 환경변수 측정장치는 기본적으로 본 실시예에 따른 멀티모달 융합을 위한 환경 변수를 측정하고, 측정된 환경 변수를 토대로 얼굴 인식, 화자 식별, 또는 이들 모두를 수행할 수 있는 기능 또는 이러한 기능을 포함한 구성부를 구비하는 장치를 지칭한다. 그리고, 환경변수 측정장치에 입력되는 입력 영상, 입력 음성, 또는 이들 모두는 입력 모달리티(modality)로 언급될 수 있다.
- [0014] 도 1을 참조하면, 본 실시예에 따른 환경변수 측정방법이 채용된 얼굴 인식 장치 또는 사용자 인식 시스템(미도시; 이하, 환경변수 측정장치라고 함)에 얼굴 인식을 위한 입력 영상(input images)이 존재하면(S110), 환경변수 측정장치는 먼저 입력 영상을 그레이 영상(gray images)으로 변환(transform)한다(S120).
- [0015] 상기 단계(S120)에서, 입력 영상을 그레이 영상으로 변환한 것은 이후의 단계에서 입력 영상에 대한 등록 영상으로부터의 거리의 변화량(variance of distances from the enrolled images)을 좀더 정확하게 구하기 위한 것이다. 다시 말해서, 그것은 등록 영상을 기준으로 입력 영상에 대한 밝기 비율이나 밝기 영역을 명확하게 구분하기 위한 것이다.
- [0016] 다음, 환경변수 측정장치는 등록 영상을 기준으로 입력 영상에 대한 영상관련 환경변수를 구한다(S130). 본 실시예에서, 입력 영상에 대한 영상관련 환경변수를 "BrightRate"이라 한다. BrightRate을 수식으로 나타내면 수학적 식 1과 같다.

수학적 식 1

[0017]
$$BrightRate = variance(distNorm(I_{enroll}, I_{test}))$$

- [0018] 수학적 식 1에서, I_{enroll} 은 등록 영상을 나타내고, I_{test} 는 테스트 영상 또는 입력 영상을 나타낸다. 수학적 식 1에 나타난 바와 같이, 본 실시예의 환경변수 측정장치에서는 등록 영상 I_{enroll} 과 테스트 영상 I_{test} 의 디스턴스 놈(distance norm)을 구하고, 구해진 디스턴스 놈 값의 변화량(variance)이 입력 영상에 대한 영상관련 환경변수 즉, BrightRate이 된다.
- [0019] 전술한 디스턴스 놈(distance norm)은 Absolute distance(1-norm distance), Euclidean distance(2-norm distance), Minkowski distance(p-norm distance), Chebyshev distance, Mahalanobis distance, Hamming distance, Lee distance, Levenshtein distance 등 모든 가능한 거리 계산 방법들 중 적어도 어느 하나에 의해 계산될 수 있다.
- [0020] 다음, 화자 식별을 위한 입력 음성(input voice)이 존재하면(S140), 환경변수 측정장치는 등록 음성을 기준으로 입력 음성에 대한 음성관련 환경변수를 구한다(S150). 본 실시예에서, 입력 음성에 대한 음성관련 환경변수를 "NoiseRate"이라 한다. NoiseRate을 수식으로 나타내면 수학적 식 2와 같다.

수학식 2

$$NoiseRate = 10 * \log \frac{\sum (x_{clean}(t))^2}{(x_{current}(t))^2}$$

[0021]

[0022] 수학식 2에서, $X_{clean}(t)$ 는 사용자를 등록(enrollment)하는 환경에서의 등록 음성 또는 목적 음성(target speech)을 나타내고, $X_{current}(t)$ 는 임의의 환경에서의 입력 음성을 나타낸다.

[0023] 상기 단계(S150)에 따르면, SNR(signal-to-noise ratio)의 측정은 어렵지만, 등록시의 입력 음성 즉 목적 음성이 순수 신호(signal)라는 가정하에서 목적 음성을 토대로 입력 음성의 환경 변수를 측정할 수 있다.

[0024] 이러한 본 실시예의 환경변수 측정방법은 예컨대 화자 식별을 위해 SNR을 이용하는 방법의 대안 방법이 될 수 있다. 즉, SNR 측정은 어떤 구간이 신호 구간이고 어떤 구간이 잡음 구간인지를 구하기 어렵기 때문에 환경의 SNR 측정으로 화자 인식이 어렵다는 단점이 있다. 하지만, 본 실시예의 NoiseRate은 등록시의 목적 음성이 순수 신호라는 가정하에서 입력 음성의 환경 변수를 측정하므로 신호 구간과 잡음 구간의 구분이 용이한 장점이 있다.

[0025] 다음, 상기 단계들(S130, S150)에서 각각 얻은 BrightRate, NoiseRate, 또는 이들 모두가 기설정된 기준값(threshold)에 미달되는지를 판단한다(S160). 기준값은 BrightRate의 경우 얼굴 인식 가능한 입력 데이터로서의 최대 임계값으로 설정될 수 있고, NoiseRate의 경우 화자 인식 가능한 입력 데이터로서의 최대 임계값으로 설정될 수 있다. 예컨대, 기준값은 사용자 식별의 한계를 고려할 때, NoiseRate의 경우 20dB 이하에서 설정될 수 있다.

[0026] 다음, 상기 단계(S160)의 판단 결과, BrightRate, NoiseRate, 또는 이들 모두가 기준값보다 크면, 해당 입력 데이터는 폐기되거나 사용할 수 없음이 사용자 등에게 통지된다(S170).

[0027] 또한, 상기 단계(S160)의 판단 결과, BrightRate, NoiseRate, 또는 이들 모두가 기준값과 같거나 작으면, 해당 입력 데이터는 얼굴 인식을 수행하는 유닛이나 화자 식별을 수행하는 유닛에 전달되어 사용자 식별용 데이터로써 사용된다(S180). 예컨대, 사용자 식별용 데이터는 정규화된 얼굴, 정규화된 음성, 또는 이들 모두에 대한 특징 추출값(feature extraction)을 포함할 수 있다.

[0028] 이와 같이, 본 실시예에 따르면, 얼굴 인식이나 화자 식별을 위한 입력 모달리티에 대한 환경 변수를 등록 모달리티에 기초하여 측정함으로써 입력 데이터에 대한 신뢰도를 신속히 판단할 수 있고, 그것에 의해 사용자 인식 시스템의 성능을 향상시킬 수 있다.

[0029] 전술한 바와 같이, 본 실시예에서는, 등록된 사용자 인식 정보를 토대로 환경 변수를 적용함으로써 효율적으로 멀티 모달 정보를 혼합하는 방법을 제공한다. 본 알고리즘의 주요한 특징은 특정 환경 조건이 특정 모달리티에 대한 낮은 정확도를 유발할 수 있고, 나머지 모달리티가 그러한 조건에 영향을 미치지 않는다는 데 기초한다. 또한, 본 실시예는 화자 식별, 얼굴 인식, 또는 이들 모두가 등록 과정(enrollment step)을 사용하는 점에 기초한다. 즉, 조합 오디오-비주얼 처리 결과로서 환경 변수에 근거하여 신뢰할 수 있는 특징을 구별적으로 선택하는 것은 본 실시예의 주된 기술적 특징 중 하나이다.

[0030] 이하에서는 전술한 실시예에 대하여 실제의 다양한 입력 영상들을 예를 들어 좀더 구체적으로 설명한다.

[0031] 도 2는 도 1의 멀티모달 융합을 위한 환경변수 측정방법에 채용가능한 등록 얼굴 영상의 일례를 나타내는 도면이다. 도 3은 도 1의 멀티모달 융합을 위한 환경변수 측정방법에 있어서 다양한 입력 영상에 대한 얼굴 인식 과정을 설명하기 위한 도면이다. 도 4는 도 3의 다양한 입력 영상에 대한 밝기를 설명하기 위한 도면이다.

[0032] 도 2, 도 3, 및 도 4에 도시한 얼굴 영상들은 예일-B 데이터베이스에서 얻은 것이다. 예일-B 데이터베이스는 여러 방향에서 조도가 변화된 얼굴 영상들을 포함한다. 또한, 예일-B 데이터베이스는 그레이 영상들을 포함한다. 도 4의 영상들(a) 내지 (f)는 도 3의 (a) 내지 (f) 라인의 왼쪽 첫번째 컬럼의 영상들에 각각 대응된다.

[0033] 도 3의 왼쪽 첫번째 컬럼에 도시된 그레이 영상들은 도 1의 2번째 단계(S120)에서의 그레이 영상에 대응할 수 있다. 도 3의 2번째와 3번째 컬럼 영상들은 도 2의 노멀 입력 영상 즉 등록 영상(200)에 대한 X축과 Y축의 상대

적인 밝기를 각각 나타낸다. 본 실시예에서는 도 2의 노멀 입력 영상을 등록 영상(200)으로 가정한다.

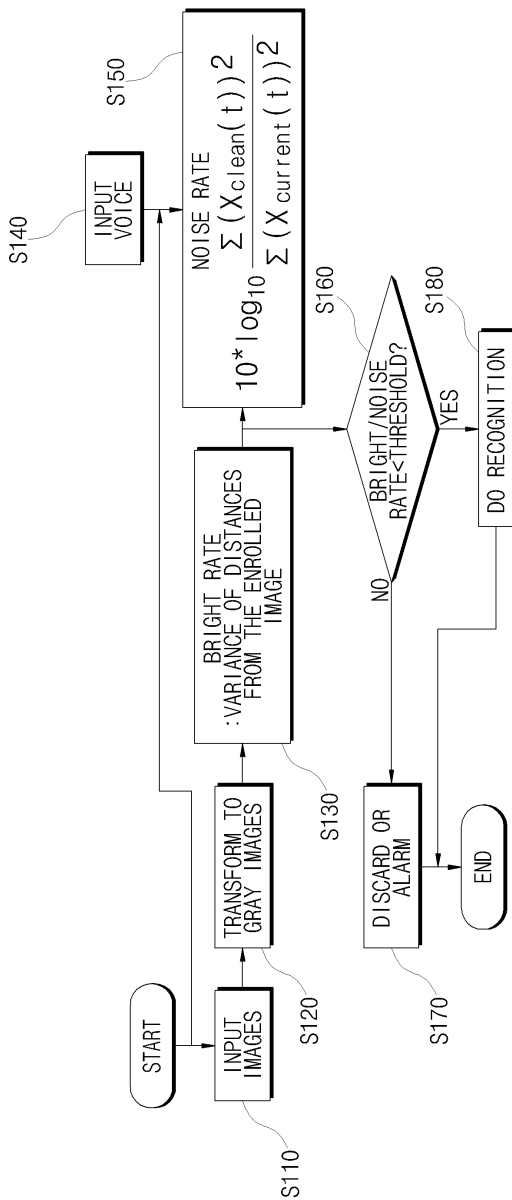
- [0034] 입력 영상의 조도가 등록 영상의 조도와 동일하거나 유사하다면, 입력 영상의 조도 라인의 기울기는 등록 영상의 조도 라인의 기울기에 근접하게 된다.
- [0035] 따라서, BrightRate이 미리 설정한 영상 인식 기준의 최대치인 기준값(threshold)보다 크면, 입력 영상은 버려지고, 새로운 입력 영상을 위해 사용자에게 빛 조건을 변화시켜 입력 영상을 준비하도록 주문 또는 요청할 수 있다.
- [0036] 도 3에 있어서, 첫번째 컬럼에서 첫번째 라인(a)의 영상은 대단히 어둡고, 따라서 코 주변의 픽셀들을 제외한 모든 픽셀은 블랙에 대단히 가깝다. 본 실시예에서 첫번째 라인(a)의 영상은 무시될 수 있다.
- [0037] 2번째 라인(b)의 영상은 거의 균일한 조도 변화를 가진다. 즉, 2번째 라인(b)의 영상은 X축과 Y축 방향에서 거의 일정한 조도 변화(uniformed illumination change)를 가진다. 따라서, 2번째 라인(b)의 영상에 대한 BrightRate 값은 상대적으로 작고, 그것에 의해 다른 영상들에 비해 해당 입력 영상의 신뢰도가 높음을 알 수 있다.
- [0038] 3번째 라인(c) 및 5번째 라인(e)의 영상들은 수직 방향의 빛 변화보다 수평 방향의 빛 변화에 의해 영향을 받는다. 따라서, 3번째 라인(c) 및 5번째 라인(e)의 영상들은 각각 수직 방향에서의 BrightRate 값보다 더 큰 수평 방향에서의 BrightRate 값을 가진다.
- [0039] 그리고, 4번째 라인(d) 및 6번째 라인(f)의 영상들은 수평 방향의 빛 변화에 의해 큰 영향을 받는다. 즉, 4번째 라인(d) 및 6번째 라인(f)의 영상들은 대응하는 3번째 라인(c) 및 5번째 라인(e)의 영상들의 수평 방향에서의 BrightRate 값보다 더 큰 수평 방향에서의 BrightRate 값을 가진다. 따라서, 4번째 라인(d) 및 6번째 라인(f)의 영상들에 대한 BrightRate 값은 3번째 라인(c) 및 5번째 라인(e)의 영상들에 대한 BrightRate 값보다 크며, 그것에 의해 4번째 라인(d) 및 6번째 라인(f)의 영상들에 대한 신뢰도는 3번째 라인(c) 및 5번째 라인(e)의 영상들에 대한 신뢰도보다 낮음을 알 수 있다.
- [0040] 이와 같이, 본 실시예에서는 등록된 영상과 테스트 영상(또는 입력 영상) 사이의 거리의 변화량으로서 신규한 개념, BrightRate를 제공한다. BrightRate은 적어도 조도에 따라서 등록된 영상을 기준으로 입력 영상의 상대적인 변화를 최대 거리로 정규화하여 표시한다. 따라서, 입력 영상의 신뢰도를 손쉽게 판단할 수 있다.
- [0041] 도 5는 도 1의 멀티모달 융합을 위한 환경변수 측정방법에 있어서 조명 거리에 따른 BrightRate을 설명하기 위한 도면이다. 도 6은 도 1의 멀티모달 융합을 위한 환경변수 측정방법에 있어서, BrightRate에 따른 인식 에러율을 나타내는 그래프이다.
- [0042] 도 5의 (a), (b), 및 (c)에 있어서, 세로축은 BrightRate, 가로축은 조명 거리(illumination distance)를 나타내며, (a)는 x-축 방향에서의 변화, (b)는 y-축 방향에서의 변화, 및 (c)는 x-축과 y-축의 두 방향 모두에서의 변화를 나타낸다.
- [0043] BrightRate는 도 5에 도시한 바와 같이 조명 거리가 약 1.5미터보다 작을 때 큰 값을 가지고, 도 6에 도시한 바와 같이, BrightRate이 높으면, 얼굴 인식에서 에러율이 높음을 알 수 있다.
- [0044] 한편, 1초에 30개 이상의 영상을 얻을 수 있고, 또한 규칙적으로 조명을 켜고 끌 수 있는 현재의 환경에서 더 이상 나쁜 조건의 입력 영상을 사용하여 얼굴 인식을 수행하는 것을 고려할 필요가 없다. 따라서, 등록 영상을 기준으로 실시간으로 입력 영상의 조도 비율이나 조도 영역의 차이 또는 변화량을 측정함으로써 손쉽게 사용자 인식을 위한 입력 데이터의 신뢰도를 판단할 수 있다.
- [0045] 전술한 실시예에 따르면, BrightRate 및 NoiseRate를 함께 채용함으로써, 주변의 잡음과 주변의 빛을 고려하는 경우에도 멀티모달 인식율을 상승시킬 수 있다.
- [0046] 위에서 설명한 바와 같이, 본 실시예에서는 얼굴 인식에 있어서 빛의 방향을 결정하거나 별도로 그림자를 보정할 필요없이 입력된 얼굴 영상을 기등록된 기준 영상의 환경 변수 기반으로 정규화함으로써 실시간으로 실제 입력된 영상 속의 노이즈 성분을 제거하고, 그것으로부터 입력 영상에 대한 얼굴 인식을 효과적으로 수행할 수 있다.
- [0047] 또한, 전술한 얼굴 인식과 유사한 방식으로 음성 인식에 있어서 입력된 음성 데이터를 기등록된 기준 데이터의 환경 변수 기반으로 정규화함으로써 실시간으로 실제 입력된 음성 속의 노이즈 성분을 제거하고, 그것으로부터 입력 음성에 대한 화자 인식을 효과적으로 수행할 수 있다.

[0048] 게다가, 전술한 얼굴 인식을 위한 환경 변수와 음성 인식을 위한 환경 변수를 융합함으로써 사용자 인식의 에러율을 현저하게 낮출 수 있다. 또한, 이러한 본 실시예의 기술에 따르면, 사용자 인식의 멀티모달 융합에 있어서 가중치나 파라미터로 사용할 수 있으며, 입력 정보의 신뢰도를 높일 수 있으므로 사용자 인식 시스템의 처리 속도나 성능을 향상시킬 수 있다.

[0049] 이상에서와 같이 상세한 설명과 도면을 통해 본 발명의 최적 실시예를 개시하였다. 용어들은 단지 본 발명을 설명하기 위한 목적에서 사용된 것이며, 의미 한정이나 특허청구범위에 기재된 본 발명의 범위를 제한하기 위하여 사용된 것은 아니다. 그러므로 본 기술 분야의 통상의 지식을 가진 자라면 본 명세서로부터 다양한 변형 및 균등한 타 실시예가 가능하다는 점을 이해할 것이다. 따라서 본 발명의 진정한 기술적 보호 범위는 첨부된 특허청구범위의 기술적 사상에 의해 정해져야 할 것이다.

도면

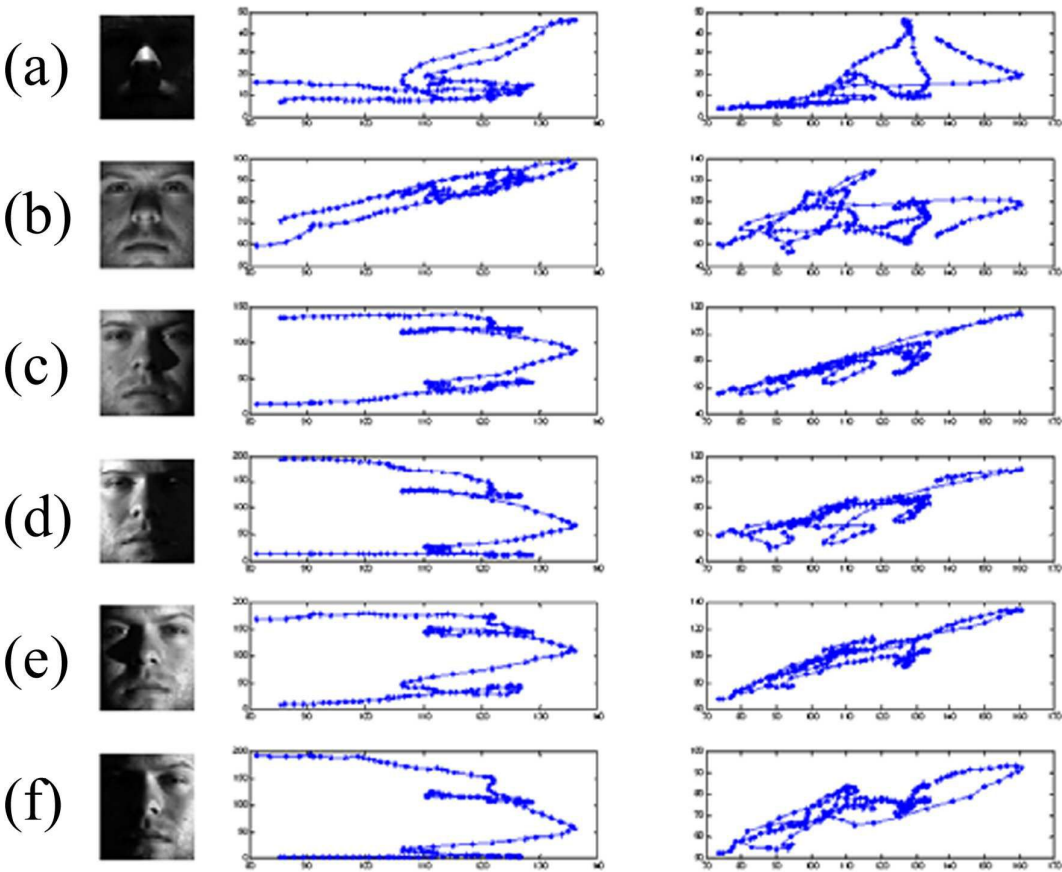
도면1



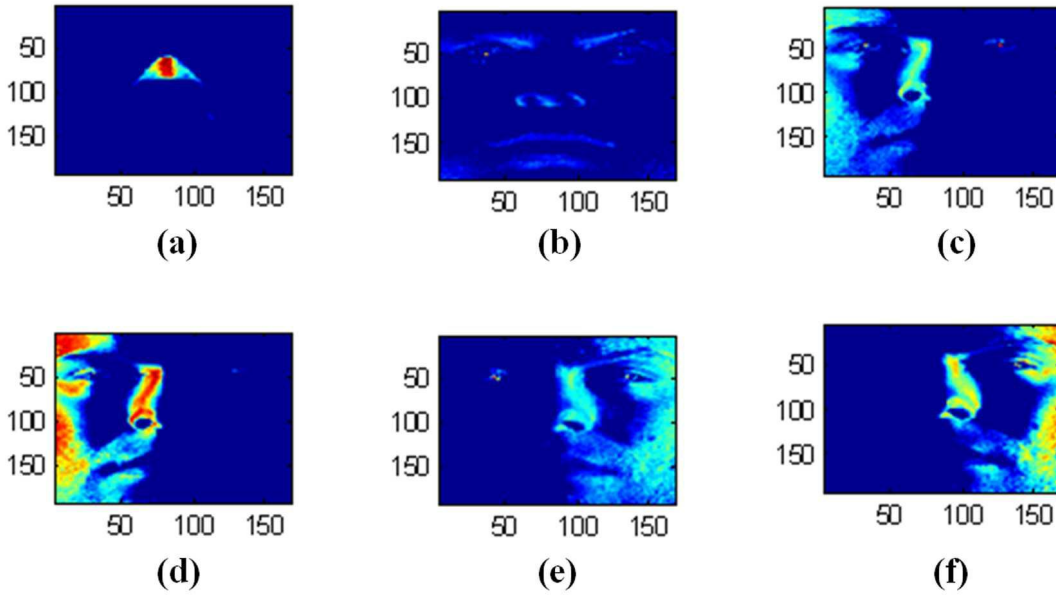
도면2



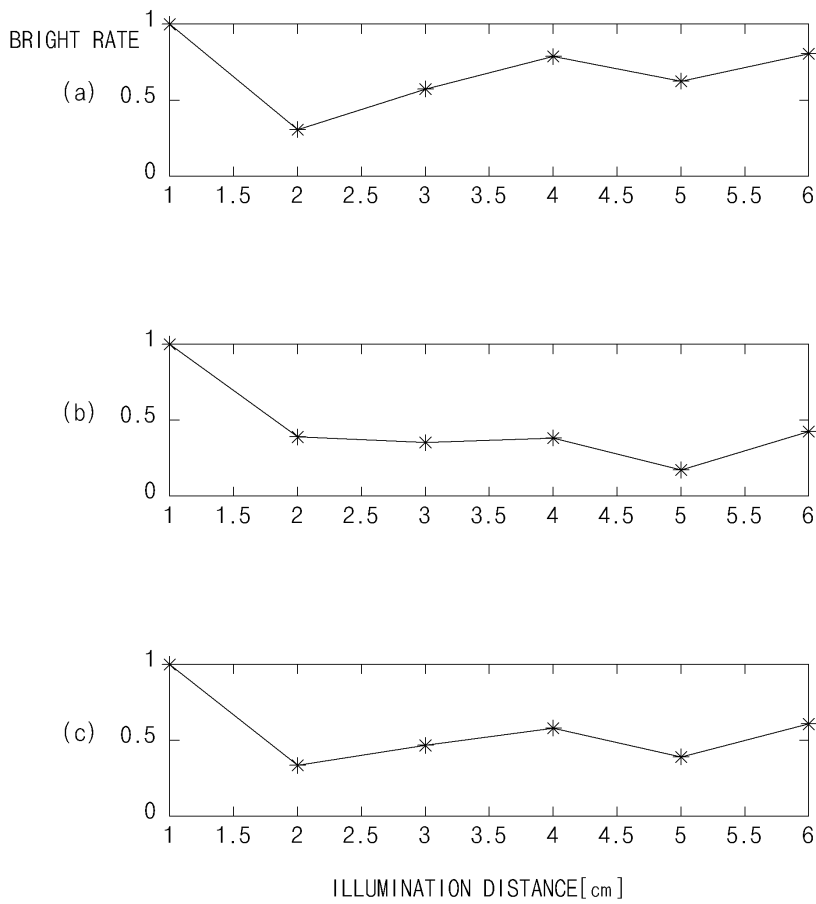
도면3



도면4



도면5



도면6

