



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2008년08월12일
(11) 등록번호 10-0851981
(24) 등록일자 2008년08월06일

(51) Int. Cl.

G06T 7/00 (2006.01) G06T 7/40 (2006.01)

G06T 7/60 (2006.01) H04N 7/24 (2006.01)

(21) 출원번호 10-2007-0015530

(22) 출원일자 2007년02월14일

심사청구일자 2007년02월14일

(56) 선행기술조사문헌

KR1020070068408 A

(뒷면에 계속)

(73) 특허권자

삼성전자주식회사

경기도 수원시 영통구 매탄동 416

(72) 발명자

박규태

경기 안양시 동안구 평촌동 932-2 꿈마을아파트 806-501

왕하이다오

중국 베이징시 하이덴구 종관쿤 난다지에 과학기술타워 4층 9호

문영수

서울 양천구 목6동 목동1단지아파트 123동 1404호

(74) 대리인

리앤목특허법인

전체 청구항 수 : 총 20 항

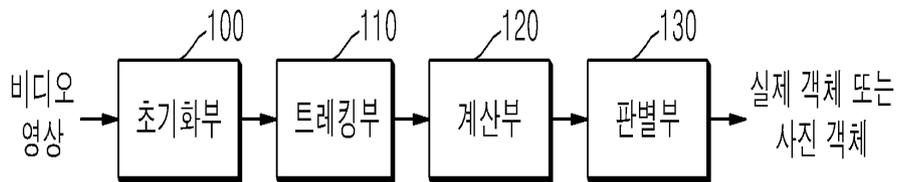
심사관 : 김혜원

(54) 비디오 영상에서 실 객체 판별 방법 및 장치

(57) 요약

본 발명에 따라 비디오 영상으로부터 객체 특징점들을 트래킹하고, 검출된 객체 특징점을 어파인 변환하고, 어파인 변환된 객체 특징점의 특성을 이용하여 비디오 영상이 실 객체 인지 여부를 판별함으로써, 사용자의 개입과 추가적인 하드웨어를 구비할 필요가 없으며, 스푸핑에 강인한 실 객체 판별(liveness detection)을 할 수 있는 효과가 있다.

대표도 - 도1



(56) 선행기술조사문헌
KR1020020037660 A
KR1020050082252 A
JP11136664 A
1019960036766
1020060101255

특허청구의 범위

청구항 1

- (a) 비디오 영상의 소정의 프레임에서 객체 특징점을 검출하여 트래킹을 초기화하는 단계;
- (b) 상기 검출된 객체 특징점을 기초로 상기 비디오 영상의 다음 프레임들에 대한 객체 특징점들을 트래킹하는 단계;
- (c) 상기 검출된 객체 특징점 및 상기 트래킹된 제 $n-1$ 프레임의 객체 특징점을 각각 제 n 프레임에 대해 어파인 변환하고, 상기 어파인 변환된 객체 특징점들과 상기 트래킹된 제 n 프레임의 객체 특징점과의 통계적 거리를 계산하는 단계; 및
- (d) 상기 계산된 거리 값을 기초로 실제 값을 계산하고, 상기 계산된 실제 값에 따라 상기 비디오 영상이 실 객체인지 여부를 판별하는 단계를 포함하는 비디오 영상에서의 실 객체 판별 방법.

청구항 2

제 1 항에 있어서,

상기 (c) 단계는,

상기 어파인 변환된 검출된 객체 특징점과 상기 제 n 프레임의 객체 특징점 간의 거리(D^1), 상기 어파인 변환된 제 $n-1$ 프레임의 객체 특징점과 상기 제 n 프레임의 객체 특징점과의 거리(D^2) 및 상기 거리(D^1)의 분산(V^1)을 각각 계산하는 것을 특징으로 하는 비디오 영상에서의 실 객체 판별 방법.

청구항 3

제 1 항에 있어서,

상기 (c) 단계는,

상기 검출된 객체 특징점, 상기 제 $n-1$ 프레임의 객체 특징점 및 상기 제 n 프레임의 객체 특징점을 정규화시키는 단계;

상기 정규화된 검출된 객체 특징점 및 제 $n-1$ 프레임의 객체 특징점들을 각각 제 n 프레임에 대해 어파인 변환을 수행하는 단계; 및

상기 어파인 변환된 검출된 객체 특징점과 상기 제 n 프레임의 객체 특징점 간의 거리(D^1), 상기 어파인 변환된 제 $n-1$ 프레임의 객체 특징점과 상기 제 n 프레임의 객체 특징점과의 거리(D^2) 및 상기 거리(D^1)의 분산(V^1)을 각각 계산하는 단계를 포함하는 것을 특징으로 하는 비디오 영상에서의 실 객체 판별 방법.

청구항 4

제 3 항에 있어서,

상기 (d) 단계에서,

상기 실제 값(L)은,

상기 분산(V^1)을 상기 거리(D^1)와 상기 거리(D^2)로 나눈 값인 것을 특징으로 하는 비디오 영상에서의 실 객체 판별 방법.

청구항 5

제 4 항에 있어서,

상기 실제 값(L)이 소정의 임계치 이상인지를 판단하는 것을 특징으로 하는 비디오 영상에서의 실 객체 판별 방법.

청구항 6

제 2 항에 있어서,

상기 (c) 단계는,

상기 어파인 변환된 제 $n-1$ 프레임의 객체 특징점과 상기 제 n 프레임의 객체 특징점과의 거리(D^2)가 소정의 임계치 이상인지를 판단하는 단계를 더 포함하고,

상기 거리(D^2)가 소정의 임계치 이상인 경우, 상기 (a) 단계를 수행하는 것을 특징으로 하는 비디오 영상에서의 실 객체 판별 방법.

청구항 7

제 1 항에 있어서,

상기 (a) 단계는,

상기 비디오 영상의 소정의 프레임에서 소정의 객체를 검출하는 단계; 및

상기 검출된 객체 영역에서 소정의 코너 점을 검출하는 단계를 더 포함하고,

상기 검출된 코너 점을 기초로 상기 특징점을 선택하는 것을 특징으로 하는 비디오 영상에서의 실 객체 판별 방법.

청구항 8

제 7 항에 있어서,

상기 객체는 얼굴인 것을 특징으로 하는 비디오 영상에서의 실 객체 판별 방법.

청구항 9

제 1 항에 있어서,

상기 (b) 단계에서,

LKT(Lucas-Kanade-Tomasi) 트래킹 알고리즘 또는 평균-시프트(mean-shift) 알고리즘을 이용하여 상기 객체 특징점들을 트래킹하는 것을 특징으로 하는 비디오 영상에서의 실 객체 판별 방법.

청구항 10

(a) 비디오 영상으로부터 객체 특징점을 검출하여 트래킹을 초기화하는 단계;

(b) 상기 객체 프레임이 검출된 프레임의 다음 프레임에 대한 객체 특징점들을 트래킹하는 단계; 및

(c) 검출된 객체 특징점과 제 $n-1$ 프레임의 객체 특징점들을 제 n 프레임에 대해 어파인 변환하고, 상기 어파인 변환된 객체 특징점들과 상기 제 n 프레임의 객체 특징점의 특성을 이용하여 상기 비디오 영상이 실 객체 인지 여부를 판별하는 단계를 포함하는 비디오 영상에서의 실 객체 판별 방법.

청구항 11

제 10 항에 있어서,

상기 (c) 단계에서,

상기 특성은,

상기 어파인 변환된 검출된 객체 특징점과 상기 제 n 프레임의 객체 특징점 간의 거리(D^1), 상기 어파인 변환된 제 $n-1$ 프레임의 객체 특징점과 상기 제 n 프레임의 객체 특징점과의 거리(D^2) 및 상기 거리(D^1)의 분산(V^1)인 것을 특징으로 하는 비디오 영상에서의 실 객체 판별 방법.

청구항 12

제 11 항에 있어서,

상기 (c) 단계에서,

상기 분산(V^1)을 상기 거리(D^1)와 상기 거리(D^2)로 나눈 값이 소정의 임계치 이상인 경우, 상기 비디오 영상을 실 객체로 판별하는 것을 특징으로 하는 비디오 영상에서의 실 객체 판별 방법.

청구항 13

제 10 항에 있어서,

상기 (b) 단계에서,

LKT(Lucas-Kanade-Tomasi) 트래킹 알고리즘 또는 평균-시프트(mean-shift) 알고리즘을 이용하여 상기 객체 특징점들을 트래킹하는 것을 특징으로 하는 비디오 영상에서의 실 객체 판별 방법.

청구항 14

제 1 항 내지 제 13 항 중 어느 한 항에 따른 방법을 컴퓨터에서 실행시키기 위한 프로그램을 기록한 기록매체.

청구항 15

비디오 영상의 소정의 프레임에서 객체 특징점을 검출하여 트래킹을 초기화하는 초기화부;

상기 검출된 객체 특징점을 기초로 상기 비디오 영상의 이후 프레임들에 대한 객체 특징점들을 트래킹하는 트래킹부;

상기 검출된 객체 특징점 및 상기 트래킹된 제 $n-1$ 프레임의 객체 특징점을 각각 어파인 변환하고, 상기 어파인 변환된 객체 특징점들과 상기 제 n 프레임의 객체 특징점과의 통계적 거리를 계산하는 계산부; 및

상기 계산된 거리 값을 기초로 실제 값을 계산하고, 상기 계산된 실제 값에 따라 상기 비디오 영상이 실 객체인지 여부를 판별하는 판별부를 포함하는 비디오 영상에서의 실 객체 판별 장치.

청구항 16

제 15 항에 있어서,

상기 계산부는,

상기 검출된 객체 특징점, 상기 제 $n-1$ 프레임의 객체 특징점 및 상기 제 n 프레임의 객체 특징점을 정규화시키는 정규화부;

상기 정규화된 검출된 객체 특징점 및 제 $n-1$ 프레임의 객체 특징점들을 각각 제 n 프레임에 대해 어파인 변환을 수행하는 어파인 변환부; 및

상기 어파인 변환된 검출된 객체 특징점과 상기 제 n 프레임의 객체 특징점 간의 거리(D^1), 상기 어파인 변환된 제 $n-1$ 프레임의 객체 특징점과 상기 제 n 프레임의 객체 특징점과의 거리(D^2) 및 상기 거리(D^1)의 분산(V^1)을 계산하는 거리 계산부를 포함하는 것을 특징으로 하는 비디오 영상에서의 실 객체 판별 장치.

청구항 17

제 16 항에 있어서,

상기 판별부는,

상기 분산(V^1)을 상기 거리(D^1)와 상기 거리(D^2)로 나눈 실제 값(L)을 계산하는 실제 값 계산부; 및

상기 계산된 실제 값(L)이 소정의 임계치 이상인지를 판단함으로써, 상기 비디오 영상이 실 객체인지를 판별하는 실 객체 판별부를 포함하는 것을 특징으로 하는 비디오 영상에서의 실 객체 판별 장치.

청구항 18

제 15 항에 있어서,

상기 초기화부는,

상기 비디오 영상의 소정의 프레임에서 소정의 객체를 검출하는 객체 검출부; 및

상기 검출된 객체 영역에서 소정의 특징점을 검출하는 특징 검출부를 포함하는 것을 특징으로 하는 비디오 영상에서의 실 객체 판별 장치.

청구항 19

제 18 항에 있어서,

상기 초기화부는,

상기 검출된 객체 영역에서 소정의 코너 점을 검출하는 코너 점 검출부를 더 포함하고,

상기 특징 검출부는,

상기 검출된 코너 점을 기초로 상기 특징점을 선택하는 것을 특징으로 하는 비디오 영상에서의 실 객체 판별 장치.

청구항 20

제 15 항에 있어서,

상기 객체는 얼굴인 것을 특징으로 하는 비디오 영상에서의 실 객체 판별 장치.

명세서

발명의 상세한 설명

발명의 목적

발명이 속하는 기술 및 그 분야의 종래기술

- <15> 본 발명의 실 객체 판별 방법 및 장치에 관한 것으로, 더 상세하게는 비디오 영상에서의 2차원 평면 객체와 3차원 비-평면 객체를 구별하여 실 객체를 판별하는 방법 및 장치에 관한 것이다.
- <16> 최근 생체 인식에 관한 많은 연구들이 진행되고 있고, 이러한 생체 인식이 얼굴 또는 홍채 인식을 이용한 생체 인증 시스템, 모바일 폰 보안, 모바일 커머스, 물리적/논리적 액세스 제어, e-커머스 등에 많이 응용되고 있지만, 실 검출(liveness detection)에는 관심을 가지지 못하고 있다.
- <17> 한편, 생체 인식 시스템은 지문, 홍채, 얼굴이 쉽게 위조될 수 있어서, 생체 시스템 보안에 있어서 위조에 취약함을 극복하는 것이 중요 이슈로 부각하고 있다. 이러한 위조를 소위 '스푸핑(spoofing)'이라고 하는데, 이는 가짜 생체 샘플을 이용하여 생체 시스템에 침입하는 일련의 프로세스를 의미한다.
- <18> 이러한 스푸핑에 대한 해결책으로 강인한 실 객체 여부를 판별하는 기술이 필수적이다. 특히 강인한 실 얼굴 판별과 추가적인 하드웨어를 사용하지 않고도, 비침입적(non-intrusive) 방법으로 구현하는 것이 필요하다.
- <19> 종래의 실 얼굴 판별 방법들은 크게 세 가지 카테고리로 분류되는데, 움직임과 3D를 이용하는 기술, 사용자 인터랙션 및 추가 하드웨어를 사용하는 방법이다.
- <20> 움직임과 3D를 이용하는 기법은 트래킹된 객체의 3D 깊이의 추정을 위해서 SFM(structure from motion method)을 이용한다. 그리고, 이러한 깊이 정보는 실제 얼굴 영상과 사진 영상을 구분하기 위해 사용된다.
- <21> 사용자 인터랙션 기법은 얼굴의 특징들을 트래킹하고 나서 특징 궤도 또는 형태 변화를 해석한다. 일반적으로, 실제 얼굴 특징 영상의 궤도와 모션 사진 영상의 궤도가 유사하기 때문에 이 방법은 정확하지 못하다. 또한, 얼굴의 특징 형태를 트래킹하는 경우, 사용자 작업이 요구되기 때문에 불편함이 있다.
- <22> 미국특허 2005/0129286호는 눈의 상태 및 위치 변화를 검사함으로써 실제 입력 여부를 추정한다. 이는 일반적인 영상 하드웨어 문제점, 초당 25 또는 30 프레임으로 영상을 촬영하기 때문에, 눈 깜빡거림이 쉽게 검출될 수

없고, 따라서 사용자는 시스템 요구 조건을 위해 초점을 신중하게 변화시켜야만 한다.

<23> 미국특허 2002/0183624호 및 2005/0134427호는 추가적인 하드웨어를 구비하는 카테고리에 속한다. 0183624호에는 실제 샘플과 위조 샘플을 구별하기 위해 능동형 발광 장치를 사용한다. 이는 알고리즘이 강인하고 사용자의 조작을 강제하지 않지만, 응용의 범위를 제한하는 특별한 발광 하드웨어가 필요하다는 단점이 있다. 마찬가지로 0134427호 또한 몸의 체온을 검사하기 위한 특별한 하드웨어를 요구한다는 문제점이 있었다.

발명이 이루고자 하는 기술적 과제

<24> 본 발명은 전술한 종래기술의 문제점을 해결하기 위해 안출된 것으로, 비디오 시퀀스 영상들을 이용하여 2차원 평면 객체와 3차원 비-평면 객체를 구별함으로써 강인한 실 객체 추정을 위해, 추가적인 하드웨어를 구비할 필요가 없으며, 사용자의 조작을 강제하지 않는, 실 객체를 판별하는 방법 및 장치를 제공하는 데 목적이 있다.

발명의 구성 및 작용

<25> 본 발명의 기술적 과제를 달성하기 위한 비디오 영상에서의 실 객체 판별 방법은 비디오 영상의 소정의 프레임에서 객체 특징점을 검출하여 트래킹을 초기화하는 단계; 상기 검출된 객체 특징점을 기초로 상기 비디오 영상의 다음 프레임들에 대한 객체 특징점들을 트래킹하는 단계; 상기 검출된 객체 특징점 및 상기 트래킹된 제 $n-1$ 프레임의 객체 특징점을 각각 제 n 프레임에 대해 어파인 변환하고, 상기 어파인 변환된 객체 특징점들과 상기 트래킹된 제 n 프레임의 객체 특징점과의 통계적 거리를 계산하는 단계; 및 상기 계산된 거리 값을 기초로 실제 값을 계산하고, 상기 계산된 실제 값에 따라 상기 비디오 영상이 실 객체인지 여부를 판별하는 단계를 포함한다.

<26> 본 발명의 다른 기술적 과제를 달성하기 위한 하는 비디오 영상에서의 실 객체 판별 방법은 비디오 영상으로부터 객체 특징점을 검출하여 트래킹을 초기화하는 단계; 상기 객체 프레임이 검출된 프레임의 다음 프레임에 대한 객체 특징점들을 트래킹하는 단계; 및 검출된 객체 특징점과 제 $n-1$ 프레임의 객체 특징점들을 제 n 프레임에 대해 어파인 변환하고, 상기 어파인 변환된 객체 특징점들과 상기 제 n 프레임의 객체 특징점의 특성을 이용하여 상기 비디오 영상이 실 객체 인지 여부를 판별하는 단계를 포함한다.

<27> 본 발명의 또 다른 기술적 과제를 달성하기 위한 비디오 영상에서의 실 객체 판별 장치는 비디오 영상의 소정의 프레임에서 객체 특징점을 검출하여 트래킹을 초기화하는 초기화부; 상기 검출된 객체 특징점을 기초로 상기 비디오 영상의 다음 프레임들에 대한 객체 특징점들을 트래킹하는 트래킹부; 상기 검출된 객체 특징점 및 상기 트래킹된 제 $n-1$ 프레임의 객체 특징점을 각각 어파인 변환하고, 상기 어파인 변환된 객체 특징점들과 상기 제 n 프레임의 객체 특징점과의 통계적 거리를 계산하는 계산부; 및 상기 계산된 거리 값을 기초로 실제 값을 계산하고, 상기 계산된 실제 값에 따라 상기 비디오 영상이 실 객체인지 여부를 판별하는 판별부를 포함한다.

<28> 본 발명의 또 다른 기술적 과제를 달성하기 위한 상기 방법을 컴퓨터에서 실행시키기 위한 프로그램을 기록한 기록매체를 포함한다.

<29> 본 발명의 세부 및 개선 사항은 종속항에 개시된다.

<30> 이하, 본 발명의 바람직한 실시 예들을 첨부한 도면들을 참조하여 상세히 설명한다.

<31> 도 1은 본 발명의 일 실시 예에 따른 실 객체 판별 장치의 개략적인 구성을 도시한 블록도이다.

<32> 도 1을 참조하면, 실 객체 판별 장치는 초기화부(100), 트래킹부(110), 계산부(120) 및 판별부(130)를 포함한다.

<33> 초기화부(100)는 입력된 비디오 영상의 특정 프레임에서 객체 특징점을 검출하여 트래킹을 초기화시킨다. 여기서, 객체 검출과 객체의 특징점 검출은 첫 번째 프레임에서 검출될 수도 있고, 이후 프레임에서 검출될 수도 있다. 비디오 영상에서 객체 검출과 객체의 특징점 검출이 완료될 때까지 반복한다.

<34> 트래킹부(110)는 초기화부(100)에서 객체와 객체의 특징점이 검출된 경우, 소정의 트래킹 알고리즘, 예를 들면 LKT 광류(optical flow)의 피라미드 알고리즘, 평균-시프트(mean-shift) 알고리즘 등을 이용하여, 검출된 객체 특징점을 기초로 이후 프레임들에 대한 객체 특징점들을 각각 트래킹한다. 트래킹 알고리즘에 대한 더 자세한 사항은 J.Y. Bouguet의 Pyramidal Implementation of the Lucas Kanade Feature Tracker, OpenCV Documentation, Microprocessor Research Labs, Intel Corp., 2000. 및 D. Comaniciu 등의 Real-Time Tracking of Non-Rigid Objects using Mean Shift, CVPR 2000.에 개시되어 있다.

- <35> 계산부(120)는 첫 번째 프레임 또는 이후 프레임에서 검출된 객체 특징점과 이를 기초로 트래킹된 n-1번째 프레임의 객체 특징점을 n번째 프레임, 즉 현재 프레임에 대해 각각 어파인 변환(affine transformation)을 수행한다.
- <36> 이렇게 어파인 변환된 초기 프레임에서 검출된 객체 특징점과 제n 프레임, 즉 현재 프레임의 객체 특징점과의 거리(D¹)를 계산하고, 이러한 거리 값에 대한 분산(V¹)을 계산한다. 또한, 제n-1 프레임, 즉 현재 프레임(제n 프레임)에 대한 이전 프레임의 객체 특징점과 제n 프레임, 즉 현재 프레임의 객체 특징점과의 거리(D²)를 계산한다. 어파인 변환과 통계적 거리, D¹, D² 및 V¹에 대한 자세한 설명은 도 3을 참조하여 후술한다.
- <37> 판별부(130)는 계산된 거리 값, D¹, D² 및 V¹을 이용하여, 실제 값(liveness value)을 계산하고, 계산된 실제 값에 따라 비디오 영상이 실제 객체인지 아니면 사진 객체인지 여부를 판별한다.
- <38> 도 2는 도 1에 도시된 실 객체 판별 장치의 보다 구체적인 구성을 도시한 블록도이다.
- <39> 도 2를 참조하면, 실 객체 판별 장치는 객체 검출부(200)와 특징 검출부(210)를 포함하는 초기화부(100), 트래킹부(110), 특징점 정규화부(220), 어파인 변환부(230)와 거리 계산부(240)를 포함하는 계산부(120) 및 실제 값 계산부(250)와 실 객체 판별부(260)를 포함하는 판별부(120)를 포함한다.
- <40> 객체 검출부(200)는 입력된 비디오 영상의 특정 프레임에서 객체를 검출한다. 본 발명의 바람직한 실시 예에서, 객체는 얼굴을 포함하며, 얼굴 검출을 위해 J. Vola의 아다부스트(adaboost) 알고리즘을 사용한다. 얼굴 검출에 관한 자세한 사항은 Paul Viola 등의 Robust Real-time Object Detection, vol. 57, p137-154, no. 2, 2004.에 개시되어 있다.
- <41> 특징 검출부(210)는 객체 검출부(200)에서 검출된 객체 영역에서 소정의 특징점들을 검출한다. 본 발명의 바람직한 실시 예에서, 객체 검출부(200)는 n-1 번째 프레임에서 소정의 객체 특징점을 검출한다. 특히, 첫 번째 프레임의 영상에서 얼굴 영상을 검출하고, 검출된 얼굴 영상 중 얼굴의 주요 특징을 반영하는 얼굴 특징점을 선택한다. 이 경우 객체는 사람의 얼굴이다. 검출된 객체 특징점의 예로는 눈 각각의 중심 위치에 해당하는 2개의 포인트, 눈썹 각각의 중심 위치에 해당하는 2개의 포인트, 코 끝의 위치에 해당하는 1개의 포인트, 입술의 양 끝에 해당하는 2개의 포인트 및 윗 입술의 끝에 해당하는 1개의 포인트 등을 객체 특징점으로 이용할 수 있다.
- <42> 객체 특징점을 검출하는 방법으로는 AAM(Active Appearance Models) 또는 ASM(Active Shape Models) 등이 있다. ASM은 눈의 위치를 검출하고, 검출된 눈의 위치를 기준으로 일반적인 모델(generic model)에 따른 다른 부위, 예를 들면 코, 입, 눈썹 등의 모양(shape) 정보를 이용하여 위치를 결정하는 모델이다. AAM은 모양 정보 뿐만 아니라, 질감 정보(texture information)를 더욱 고려하여 얼굴 특징을 효율적으로 설명할 수 있는 특징점을 검출하는 모델이다.
- <43> 또한, 본 발명의 바람직한 실시 예에서, 좀더 양호한 특징점을 선택하기 위해 코너(corner)를 검출할 수 있다. 이는 특징 검출을 에러를 줄이기 위해 검출된 얼굴 영역에서 모서리 부분을 검출하여 특징점을 선택하는 것이다.
- <44> 코너 검출은 다음 수학식 1 및 2를 이용하여, 영상 I의 모든 픽셀에서 G 행렬과 행렬 G의 고유값(λ_m)을 계산한다.

수학식 1

$$G = \sum_{x=p_x-w_x}^{p_x+w_x} \sum_{y=p_y-w_y}^{p_y+w_y} \begin{bmatrix} I_x^2(x,y) & I_x(x,y)I_y(x,y) \\ I_x(x,y)I_y(x,y) & I_y^2(x,y) \end{bmatrix}$$

<45>

수학식 2

$$I_x = \frac{I(x+1,y) - I(x-1,y)}{2} \quad I_y = \frac{I(x,y+1) - I(x,y-1)}{2}$$

<46>

- <47> 여기서, $I(x,y)$ 는 그레이 영상이고, (w_x, w_y) 는 영상 영역의 폭(width)과 높이(height)이다.
- <48> 행렬 G 와 행렬 G 의 고유값(λ_m)을 구하고, 특징점을 선택한다. 전체 영상에 대해 λ_m 의 최대값을 λ_{max} 라하고, $\lambda_m > 0.2\lambda_{max}$ 인 화소에 대해, 최대 화소를 찾는다. 여기서, 선택된 화소의 λ_m 값은 이웃하는 다른 화소들보다 더 크다. 또한, 선택된 화소들의 거리를 소정 값 이하로 유지한다.
- <49> 코너 검출에 관한 자세한 사항은 Jianbo Shi 등의 "Good features to track", Proc. IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recogn., p593-600, 1994에 개시되어 있다.
- <50> 트래킹부(110)는 특징 검출부(210)에서 검출된 객체의 특징점을 기초로, 다음 프레임들 각각에 대한 특징점들을 트래킹한다.
- <51> 본 발명의 바람직한 실시 예에서, 트래킹부(110)는 초기화부(100)를 통해 첫번째 프레임에서 얼굴 특징점이 검출되면, 트래킹부(110)는 두 번째 프레임과 첫 번째 프레임의 화소 값 차이를 최소로 하는 광류(optical flow)를 계산한다. 예를 들면, 첫 번째 프레임에서 검출된 얼굴 특징점 각각을 중심으로 하는 소정 크기의 템플릿 이미지의 그레이 레벨(grey level)과 상기 템플릿에 대응되는 두 번째 프레임의 템플릿 이미지의 그레이 레벨을 구하고, 템플릿 이미지에 속하는 각 픽셀이 갖는 그레이 레벨의 차를 이용하여 광류를 계산할 수 있다. 본 발명의 바람직한 실시 예에서는 광류를 계산하는 방법으로 LKT(Lucas-Kanade-Tomasi) 알고리즘을 이용하였으며, 하기 수학적식 3 내지 7에 따라 $\varepsilon(v)$ 값을 최소로 하는 광류(v) 값을 계산할 수 있다.

수학적식 3

<52>
$$\varepsilon(v) = \varepsilon(v_x, v_y) = \sum_{x=u_x-\omega_x}^{u_x+\omega_x} \sum_{y=u_y-\omega_y}^{u_y+\omega_y} (I(x,y) - J(x+v_x, y+v_y))^2$$

수학적식 4

<53>
$$I_x(x,y) = \frac{I(x+1,y) - I(x-1,y)}{2}, \quad I_y(x,y) = \frac{I(x,y+1) - I(x,y-1)}{2}$$

수학적식 5

<54>
$$G(u_x, u_y) = \sum_{x=u_x-\omega_x}^{u_x+\omega_x} \sum_{y=u_y-\omega_y}^{u_y+\omega_y} \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}, \quad \mathcal{D}(x,y) = I(x,y) - J(x,y)$$

수학적식 6

<55>
$$\bar{b} = \sum_{x=u_x-\omega_x}^{u_x+\omega_x} \sum_{y=u_y-\omega_y}^{u_y+\omega_y} \begin{bmatrix} \delta I_x \\ \delta I_y \end{bmatrix}$$

수학적식 7

<56>
$$v = G^{-1} \bar{b}$$

- <57> 여기에서, I 는 이전 ($n-1$ 번째) 프레임이고, J 는 현재 (n 번째) 프레임이며, $u = (u_x, u_y)$ 는 이전 프레임에서 하나의 특징점이고, $v = (v_x, v_y)$ 는 광류이며, (w_x, w_y) 는 광류를 계산하는 이미지의 템플릿 사이즈이고, I_x 와 I_y 는 I 의 미분 영상값이다.
- <58> 특징점 정규화부(220)는 특징점 검출부(210)에서 검출된 특징점, 즉 초기 프레임에서의 특징점, 현재 프레임(n

번째)의 특징점 및 이전 프레임(n-1 번째)의 특징점을 정규화시킨다.

<59> 특징점은 다음 수학적 식 8을 이용하여 정규화한다.

수학적 식 8

$$P_i(n) = \frac{P_i(n) - \bar{P}(n)}{\sigma_{P(n)}}$$

$$\left\{ \begin{array}{l} \bar{P}(n) = \frac{1}{N} \sum_{i=1}^N P_i(n) \\ \sigma_{P(n)} = \sqrt{\left(\frac{1}{N} \sum_{i=1}^N P_i(n) \cdot P_i(n) - \bar{P}(n) \cdot \bar{P}(n) \right)} \end{array} \right.$$

<60>

<61> 여기서, $P_i(n)$, $n=1, 2, \dots, N$ 이다.

<62> 어파인 변환부(230)는 특징점 정규화부(220)에서 정규화된 특징점들의 어파인 변환을 수행한다. 즉, 어파인 변환부(230)는 초기 프레임에서의 특징점을 현재 프레임(n 번째 프레임)에 대해 어파인 변환하고, 이전 프레임(n-1 번째 프레임)에서의 특징점을 현재 프레임(n 번째 프레임)에 대해 어파인 변환을 수행한다.

<63> T는 n-1번째 프레임의 좌표 행렬 X를 입력으로 하는 어파인 변환이고, X'는 n 번째 프레임의 좌표 행렬이다.

수학적 식 9

$$S_x = \frac{1}{n} \sum_{i=1}^n x_i \quad S_y = \frac{1}{n} \sum_{i=1}^n y_i$$

$$S_{x'} = \frac{1}{n} \sum_{i=1}^n x'_i \quad S_{y'} = \frac{1}{n} \sum_{i=1}^n y'_i$$

$$S_{xx} = \frac{1}{n} \sum_{i=1}^n x_i^2 \quad S_{yy} = \frac{1}{n} \sum_{i=1}^n y_i y_i$$

$$S_{xy} = \frac{1}{n} \sum_{i=1}^n x_i y_i$$

$$S_{xx'} = \frac{1}{n} \sum_{i=1}^n x_i x'_i \quad S_{yy'} = \frac{1}{n} \sum_{i=1}^n y_i y'_i$$

$$S_{xy'} = \frac{1}{n} \sum_{i=1}^n x_i y'_i \quad S_{yx'} = \frac{1}{n} \sum_{i=1}^n y_i x'_i$$

<64>

<65> 여기에서, (x_i, y_i) 는 이전(n-1번째) 프레임에서 결정된 객체 특징점들의 좌표이고, (x'_i, y'_i) 는 현재(n 번째) 프레임에서 결정된 객체 특징점들의 좌표이다.

수학적 식 10

$$T \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix}$$

<66>

수학적 식 11

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \frac{1}{\Delta} \begin{pmatrix} S_{xx'} & S_{yy'} \\ S_{xy'} & S_{yx'} \end{pmatrix} \begin{pmatrix} S_{yy} & -S_{xy} \\ -S_{xy} & S_{xx} \end{pmatrix}$$

$$\Delta = S_{xx'} S_{yy'} - S_{xy}^2$$

<67>

$$t_x = S_{x'} \quad t_y = S_{y'}$$

<68>

- <69> 어파인 변환 파라미터 T는 상기 수학식10과 같이 6개(t_x, t_y, a, b, c 및 d)의 요소(element)들에 의하여 특정된다.
- <70> 사진 얼굴 영상과 실제 얼굴 영상과의 가장 큰 차이점은 사진 얼굴은 깊이 변화를 가지지 않는 2차원 평면 객체이고, 실제 얼굴은 비교적 큰 깊이 변화를 가지는 3차원 객체라는 것이다.
- <71> 이상적인 영상 조건, 이상적인 정사영(orthography), 영상 잡음 및 왜곡이 없는 등의 조건과 트래킹 오류 및 바이어스가 없는 이상적인 트래킹 알고리즘에서, 서로 다른 프레임들에서 어파인 변환된 트래킹 포인트들 사이의 거리는 평면 객체, 즉 사진 얼굴에 대해서 제로이다. 반면에, 이러한 거리는 사진 얼굴 객체에서 보다 실제 얼굴 객체에 대해서 훨씬 크다.
- <72> 그러나, 사진 얼굴의 거리는 실제 경우에 있어서 명백히 제로가 아니다. 보다 강인한 실 얼굴 판별이 트래킹 효과들과 영상 조건들에 의한 외부적 요인을 배제하고, 사진 얼굴과 실제 얼굴 사이의 3차원 깊이 차이의 내재적 요인을 반영하도록 정의되어야 한다.
- <73> 어파인 변환을 설명하기 위해 도 3을 참조하면, 먼저, 제1 프레임의 특징점 i 를 $P_i(1)$ 라고 가정하고, 여기서, $i=1, 2, \dots, N$ 이고, $N>3$ 이라고 가정한다.
- <74> 제1, 2, ..., $n-1, n$ 프레임이 도시되어 있고, 제1 프레임의 특징점은 $P_i(1)$ (300), 제 $n-1$ 프레임의 특징점은 $P_i(n-1)$ (310), 제 n 프레임의 특징점은 $P_i(n)$ (320)으로 도시되어 있다.
- <75> $N>3$ 에 대해서, 첫 번째 프레임의 특징점(300)을 현재 프레임(제 n 프레임)에 대해 어파인 변환을 수행한다. 어파인 변환된 제1 프레임의 특징점은 $\tilde{P}_i(1)$ (305)이다.
- <76> 이전 프레임의 특징점(310)을 현재 프레임(제 n 프레임)에 대해 어파인 변환을 수행한다. 어파인 변환된 제 $n-1$ 프레임의 특징점은 $\tilde{P}_i(n-1)$ (315)이다.
- <77> 다시 도 2를 참조하면, 거리 계산부(240)는 어파인 변환된 제1 프레임의 특징점은 $\tilde{P}_i(1)$ (305)과 현재 프레임의 특징점 $P_i(n)$ 과의 거리, $D_i^1(n)$ 를 다음 수학식 12에 의해 계산한다.

수학식 12

<78>
$$D_i^1(n) = \|P_i(n) - \tilde{P}_i(1)\|$$

<79> 그리고, 다음 수학식 13에 의해 $D^1(n)$ 을 구한다.

수학식 13

<80>
$$D^1(n) = \frac{1}{K} \sum_{i=1}^K D_i^1(n)$$

<81> 여기서, $D^1(n)$ 은 $D_i^1(n)$ 의 평균이고, K 는 특징점의 개수이다.

<82> D_i^1 은 평면 객체들에 대해서는 제로이다. $D^1(n)$ 의 실제 값은 누적 트래킹 오차와 주로 관련된다. 이러한 $D^1(n)$ 의 실제 값은 트래킹된 포인트가 처음 위치에서 멀리 떨어진 경우에 증가한다.

<83> 또한, 어파인 변환된 제 $n-1$ 프레임의 특징점 $\tilde{P}_i(n-1)$ (315)과 현재 프레임의 특징점 $P_i(n)$ 과의 거리, $D_i^2(n)$ 를 다음 수학식 14에 의해 계산한다.

수학식 14

<84>
$$D_i^2(n) = \|P_i(n) - \tilde{P}_i(n-1)\|$$

<85> 그리고, 다음 수학식 15에 의해 $D^2(n)$ 을 구한다.

수학식 15

<86>
$$D^2(n) = \frac{1}{K} \sum_{i=1}^K D_i^2(n)$$

<87> 여기서, $D^2(n)$ 은 $D_i^2(n)$ 의 평균이고, K는 특징점의 개수이다.

<88> $D^2(n)$ 의 정의에 따르면 $D^2(n)$ 의 값은 연속한 2개의 프레임들의 특징점들의 거리를 나타낸다. 이 값은 주로 움직임 정보에 관련된다.

<89> 특정 객체가 더 빠르게 움직인다면, 움직임 흐릿함(motion blur), 트래킹 알고리즘의 탐색 범위의 초과와 같은 요인들이 발생하는데, 이러한 요인들은 트래킹 에러 또는 트래킹 바이어스의 확률을 증가시킨다. 따라서, $D^2(n)$ 의 값은 이러한 경우에 증가한다.

<90> 만약 $D^2(n)$ 의 값이 소정의 임계치(T_1)보다 크다면, 즉 $D^2 > T_1$ 이라면 심각한 트래킹 에러가 존재하는 것이므로 자동적으로 시스템을 재시작한다. 이는 급격한 움직임으로 인한 트래킹 에러를 피하기 위함이다.

<91> $D^1(n)$ 과 $D^2(n)$ 의 차이점은 $D^1(n)$ 은 누적된 트래킹 결과로부터 기인한 에러를 나타내고, $D^2(n)$ 은 순간적인 트래킹 에러를 나타낸다는 것이다.

<92> 또한, D^1 의 분산(variance)을 다음 수학식 16에 의해 계산한다.

수학식 16

<93>
$$V^1(n) = \frac{1}{K} \sum_{i=1}^K D_i^1(n) \cdot D_i^1(n) - \bar{D}^1(n) \cdot \bar{D}^1(n)$$

<94> 여기서, $V^1(n)$ 은 $D_i^1(n)$ 의 분산이고, K는 특징점의 개수이고, $\bar{D}^1(n)$ 은 $D_i^1(n)$ 의 평균이다.

<95> $V^1(n)$ 은 $D^1(n)$ 의 분산이기 때문에, 영상 왜곡과 영상 양자화, 평균 트래킹 효과들과 같은 항상 존재하는 외부적인 요인들을 감소시킬 수 있다.

<96> 만약 입력 객체가 평면이라면, $V^1(n)$ 은 작은 값이 될 것이다. 반면에 3차원 객체인 경우, 예를 들면 실제 얼굴인 경우에 $V^1(n)$ 값은 평면 객체보다 훨씬 더 클 것이다.

<97> 실제 값(liveness value) 계산부(250)는 거리 계산부(240)에서 계산된 D^1 , D^2 및 V^1 의 값을 이용하여 다음 수학식 17에 따라 실제 값을 계산한다.

수학식 17

<98>
$$L(n) = \frac{V^1(n)}{D^1(n) \cdot D^2(n)}$$

<99> 즉, 실제 값, L(n)은 $V^1(n)$ 을 $D^1(n)$ 과 $D^2(n)$ 으로 나눈 것이다.

<100> 또한, 다음 수학식 18에 따라 실제 값을 계산할 수 있다. 이는 실제 값의 강인성을 증가시키기 위한 것으로 상

기 수학적 식 17에 의해 계산된 $L(n)$ 을 F 의 프레임들로 나눈 것이다.

수학적 식 18

$$L = \frac{1}{F} \sum_{i=n}^F L(n)$$

<101>

<102> 실 객체 판별부(260)는 상기 수학적 식 18에 의해 계산된 실제 값(L)이 소정의 임계치(T_2)보다 크다면($L > T_2$), 입력된 비디오 영상을 실제 객체로 판단하고, 실제 값이 소정의 임계치(T_2)보다 작다면($L < T_2$), 입력된 비디오 영상을 사진 객체로 판단한다. 여기서 임계치(T_2)는 사용자의 설정에 의해 또는 시스템의 특성에 따라 결정될 수 있는 값이다.

<103>

요약하면, 사진 얼굴과 실제 얼굴의 영상의 가장 큰 차이점은 사진 얼굴은 평면 객체이고, 실제 얼굴은 비교적 큰 깊이 변화를 가진 3차원 객체라는 것이다. 이러한 차이점을 이용한 3가지 통계적 거리, D^1 , D^2 및 V^1 의 정의한다.

<104>

D^1 은 누적 트래킹 오차를 나타낸다. D^1 의 값은 특징점들이 초기 형태에서 벗어나는 경우 증가한다. D^2 는 주로 움직임 정보에 관계된다. 특징점들이 더 빨리 움직일수록, D^2 의 값은 더 커진다.

<105>

V^1 은 3차원 깊이 변화를 포함한다. 트래킹된 특징점들이 같은 평면에 있거나 또는 거의 깊이 변화가 없다면, V^1 의 값은 매우 작지만, 그렇지 않다면 크다.

<106>

본 발명의 바람직한 실시 예에서 정의된 실제 값, L 은 V^1 을 D^1 및 D^2 로 나눈 값인데, 이는 누적 트래킹 오차와 움직임 트래킹 오차가 보상된 트래킹된 객체의 3차원 정보를 반영한다.

<107>

도 4 내지 6은 본 발명의 또 다른 실시 예에 따른 통계적 거리, D^1 , D^2 및 V^1 의 실제 객체와 사진 객체를 비교하기 위한 도면이다.

<108>

도 4 내지 6은 실제 얼굴과 사진 얼굴에 대한 통계적 거리 D^1 , D^2 및 V^1 의 히스토그램을 나타낸다.

<109>

X축은 소정의 범위, 최소값과 최대값 사이에서, 256개로 양자화된 D^1 , D^2 및 V^1 의 값이고, Y축은 프레임 수이다.

<110>

D^1 및 D^2 의 값은 비디오 영상이 실제 얼굴(400, 500)인 경우와, 사진 얼굴(410, 510)인 경우가 일반적인 트래킹 조건 하에서 유사하게 나타난다.

<111>

하지만, V^1 은 실제 얼굴(600)인 경우에는 크게 나타나지만, 사진 얼굴(610)인 경우에는 더 작게 나타난다.

<112>

도 7은 본 발명의 또 다른 실시 예에 따른 실 객체 판별 방법을 설명하기 위한 흐름도이다.

<113>

도 7을 참조하면, 실 객체 판별 장치는 단계 700 및 702에서, 영상을 획득하고, 현재 시스템이 트래킹 모드인지 판단한다. 여기서 트래킹 모드는 비디오 영상에서 객체가 검출되고, 객체의 특징점이 검출되고 난 후에 수행되는 모드이다.

<114>

트래킹 모드가 아닌 경우, 단계 704로 진행하여 다음 프레임에 대해 객체와 특징점 검출을 수행한다. 객체와 특징점이 검출된 경우, 단계 706으로 진행하여 이후 프레임에 대한 특징점 트래킹을 수행한다.

<115>

단계 708에서, 검출된 특징점과 $n-1$ 번째 프레임의 특징점을 n 프레임에 대해 어파인 변환한다. 단계 710에서, 초기에 검출되어 어파인 변환된 특징점과 n 프레임의 특징점 사이의 거리(D^1)를 계산하고, 어파인 변환된 $n-1$ 번째 프레임의 특징점과 현재 특징점 사이의 거리(D^2)를 계산한다. 또한, 초기에 검출되어 어파인 변환된 특징점과 n 프레임의 특징점 사이의 거리(D^1)의 분산(V^1)을 계산한다.

<116>

단계 712에서, 단계 710에서 계산된 통계적 거리값, D^1 , D^2 및 V^1 을 기초로, 상기 수학적 식 18을 이용하여 실제

값(L)을 계산한다.

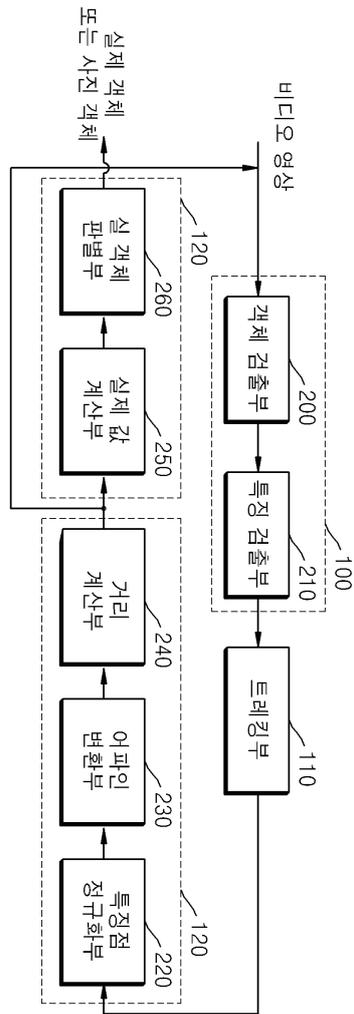
- <117> 단계 714에서, 단계 712에서 계산된 실제 값 L을 이용하여, 실제 객체 여부를 판단한다.
- <118> 도 8은 본 발명의 또 다른 실시 예에 따른 실 객체 판별 방법을 설명하기 위한 흐름도이다.
- <119> 단계 800 및 802에서, 입력된 비디오 영상에서 소정의 영상 프레임을 획득하고, 획득한 영상에서 얼굴을 검출한다. 단계 804에서, 검출된 얼굴 영역에서 코너점들을 검출한다. 단계 806에서, 검출된 코너점을 기초로 특징점을 선택한다. 단계 808에서, 트래킹 알고리즘, 예를 들면 LKT알고리즘을 이용하여 이후 프레임들에 대한 특징점들을 트래킹한다. 단계 810에서, 검출된 특징점과 트래킹된 특징점들에 대한 통계적 거리, D^1 , D^2 및 V^1 를 계산한다. 단계 812에서, D^2 가 소정의 임계치(T_1) 이상인지를 판단한다. 이는 과도한 트래킹 오차가 발생한 경우 시스템을 재시작하기 위함이다. 즉 $D^2 > T_1$ 인 경우, 다시 처음으로 돌아가 비디오 영상을 다시 획득하여 다음 프로세스들을 수행한다.
- <120> $D^2 < T_1$ 인 경우, 단계 814에서, 수학적 식 18을 이용하여 실제 값 L을 계산한다. 단계 816에서, 실제 값 L이 소정의 임계치(T_2) 이상인지를 판단하여, $L > T_2$ 인 경우 입력된 비디오 영상이 실 얼굴이라고 판단하고, $L < T_2$ 인 경우에는 사진 얼굴이라고 판단한다.
- <121> 도 9는 본 발명의 또 다른 실시 예에 따른 실 객체 판별 장치를 이용한 접근 제어시스템의 개략적인 구성을 도시한 블록도이다.
- <122> 도 9를 참조하면, 접근 제어 시스템은 카메라(800), 영상 획득부(810), 영상 제어부(820), 얼굴 검출부(830), 얼굴 인식부(840), 접근 제어부(850), DB(860), 실 얼굴 판별 장치(870), 입력부(880) 및 도어 제어부(890)를 포함한다.
- <123> 접근 제어시스템은 카메라를 통해 촬영한 비디오 영상, 예를 들면 사람의 얼굴을 검출하고, 검출된 얼굴을 인식하여 도어를 개폐하는 것이다. 여기서 본 발명의 바람직한 실시 예에 따른 실 얼굴 판별 장치를 이용하여 촬영한 비디오 영상이 사람의 실제 얼굴인지 아니면 사진 얼굴인지를 판별할 수 있다.
- <124> 한편 본 발명은 컴퓨터로 읽을 수 있는 기록 매체에 컴퓨터가 읽을 수 있는 코드로 구현하는 것이 가능하다. 컴퓨터가 읽을 수 있는 기록 매체는 컴퓨터 시스템에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록 장치를 포함한다.
- <125> 컴퓨터가 읽을 수 있는 기록 매체의 예로는 ROM, RAM, CD-ROM, 자기 테이프, 플로피디스크, 광 데이터 저장장치 등이 있으며, 또한 캐리어 웨이브(예를 들어 인터넷을 통한 전송)의 형태로 구현하는 것을 포함한다. 또한, 컴퓨터가 읽을 수 있는 기록 매체는 네트워크로 연결된 컴퓨터 시스템에 분산되어, 분산 방식으로 컴퓨터가 읽을 수 있는 코드가 저장되고 실행될 수 있다. 그리고, 본 발명을 구현하기 위한 기능적인(functional) 프로그램, 코드 및 코드 세그먼트 들은 본 발명이 속하는 기술 분야의 프로그래머들에 의하여 용이하게 추론될 수 있다.
- <126> 이제까지 본 발명에 대하여 바람직한 실시예를 중심으로 살펴보았다. 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자는 본 발명의 본질적인 특성에서 벗어나지 않는 범위에서 변형된 형태로 본 발명을 구현할 수 있음을 이해할 것이다. 그러므로, 상기 개시된 실시예 들은 한정적인 관점이 아니라 설명적인 관점에서 고려되어야 한다. 본 발명의 범위는 전술한 설명이 아니라 특허청구범위에 나타나 있으며, 그와 동등한 범위 내에 있는 모든 차이점은 본 발명에 포함된 것으로 해석되어야 한다.

발명의 효과

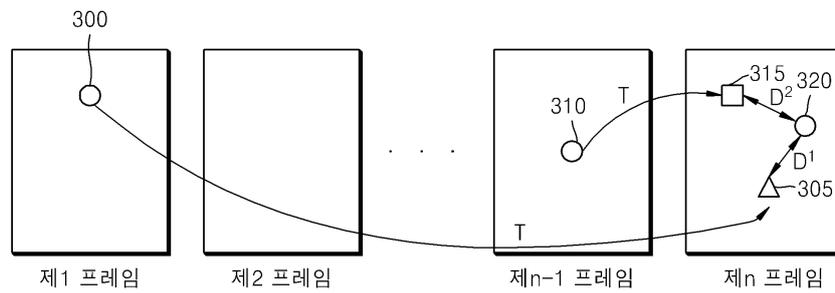
- <127> 본 발명에 따라 비디오 영상으로부터 객체 특징점들을 트래킹하고, 검출된 객체 특징점을 어파인 변환하고, 어파인 변환된 객체 특징점의 특성을 이용하여 비디오 영상이 실 객체 인지 여부를 판별함으로써, 사용자의 개입과 추가적인 하드웨어를 구비할 필요가 없으며, 스푸핑에 강한 실 객체 판별(liveness detection)을 할 수 있는 효과가 있다.
- <128> 또한, 2D 영상 처리 알고리즘만으로 구현 가능하므로, 복잡한 계산과정을 요구하지 않는 장점이 있다.

도면의 간단한 설명

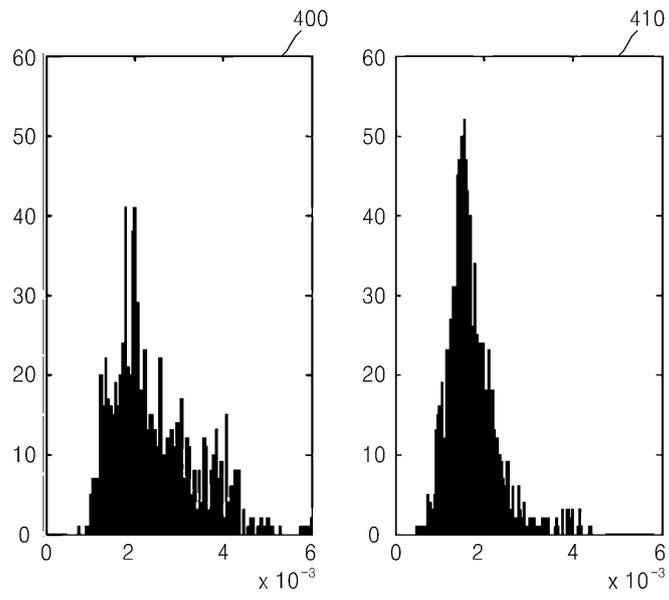
도면2



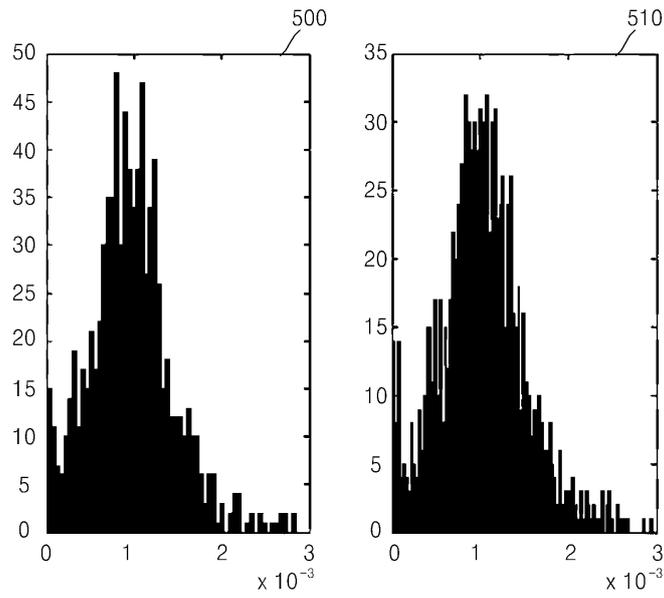
도면3



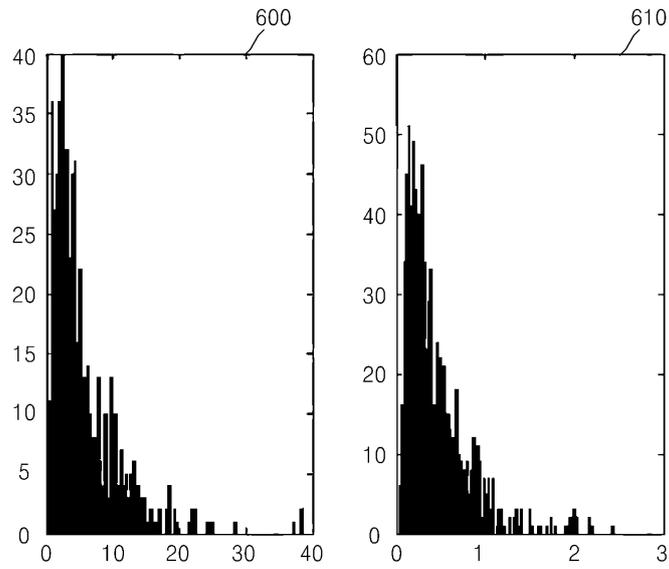
도면4



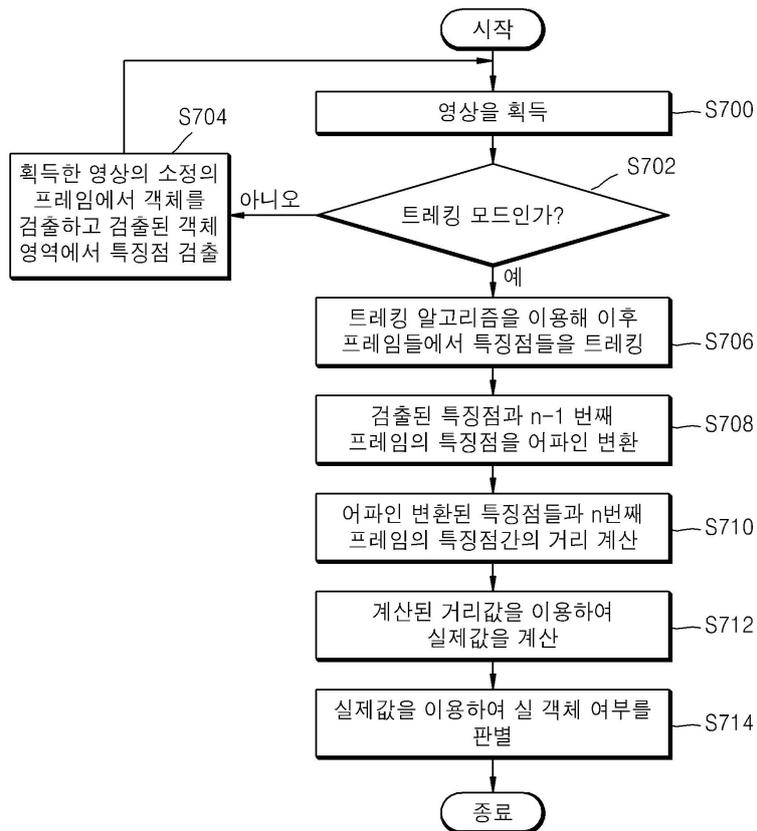
도면5



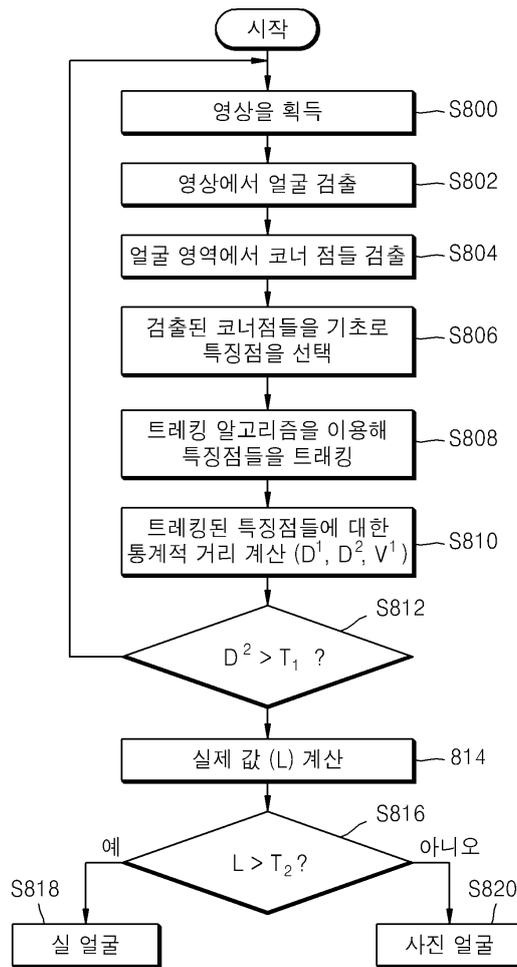
도면6



도면7



도면8



도면9

