(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2022년02월04일
(11) 등록번호 10-2358813
(24) 등록일자 2022년01월28일

(51) 국제특허분류(Int. Cl.)
G06K 9/00 (2022.01)
(52) CPC특허분류
G06V 20/00 (2022.01)
G06V 20/54 (2022.01)
(21) 출원번호 10-2015-7009258
(22) 출원일자(국제) 2013년09월12일
심사청구일자 2018년08월16일
(85) 번역문제출일자 2015년04월10일
(65) 공개번호 10-2015-0067193
(43) 공개일자 2015년06월17일
(86) 국제출원번호 PCT/US2013/059471
(87) 국제공개번호 WO 2014/043353
국제공개일자 2014년03월20일
(30) 우선권주장
61/700,033 2012년09월12일 미국(US)
13/838,511 2013년03월15일 미국(US)
(56) 선행기술조사문헌
US20080181453 A1*
US20120027299 A1*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
아비질론 포트리스 코퍼레이션
캐나다, 브이6씨 0에이3 브리티시 컬럼비아, 밴쿠버, 버라드 스트리트 2900-550
(72) 발명자
장, 종
미국 20191 버지니아주 레스톤 선라이즈 밸리 드라이브 #290 11600 오브젝트비디오 인크.
인, 웨이홍
미국 20191 버지니아주 레스톤 선라이즈 밸리 드라이브 #290 11600 오브젝트비디오 인크.
베네티아너, 피터
미국 22101 버지니아주 맥린 아이비 힐 드라이브 6623
(74) 대리인
강명구

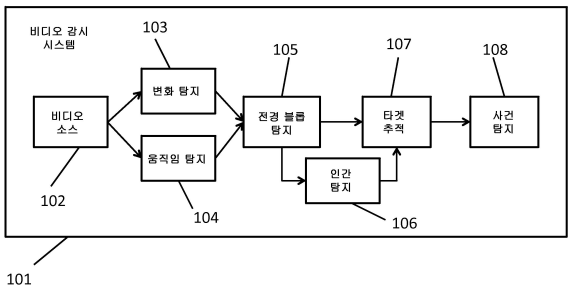
전체 청구항 수 : 총 32 항 심사관 : 노용완

(54) 발명의 명칭 **비디오 내의 객체들을 탐지하기 위한 방법들, 장치들 및 시스템들**

(57) 요 약

비디오 이미지 내의 관심의 대상인 인간들이나 다른 객체들을 탐지하도록 비디오 내용 분석을 수행하기 위한 방법들 장치들 및 시스템들이 개시된다. 인간들의 탐지는 각 인간의 위치를 결정하고 및/또는 모니터된 영역들의 군중 분석들을 수행하도록 인간들의 숫자를 계산하도록 사용될 수 있다.

대 표 도 - 도1



(52) CPC특허분류
G06V 40/103 (2022.01)

명 세 서

청구범위

청구항 1

비디오 내의 인간 객체(human object)들을 탐지하는 방법에 있어서,

비디오 이미지의 픽셀들을 전경 픽셀(foreground pixel)들로 결정하는 단계를 포함하고, 상기 전경 픽셀들의 그룹은 하나 또는 그 이상의 전경 블롭(foreground blob)들의 전경 블롭 세트를 구성하며;

상기 비디오 이미지 내의 N(여기서 N은 1보다 큰 정수이다)의 소정의 위치들의 대응되는 위치들에서 각각의 N의 소정의 형상들에 대해, 상기 대응되는 소정의 위치에서 인간의 대응되는 확률을 수득하도록 상기 대응되는 소정의 형상을 상기 전경 블롭 세트와 비교하는 단계를 포함하고, 이에 따라 상기 N의 소정의 위치들에 대응되는 N의 확률들을 수득하며;

상기 N의 확률들을 이용하여, 상기 전경 블롭 세트에 의해 나타나는 X(여기서 X는 전체 숫자이다)의 인간들을 결정하는 단계를 포함하며;

상기 X의 인간들의 표시의 결정을 이용하여 보고, 경보 및 사건 검출의 적어도 하나를 제공하는 단계를 포함하는 것을 특징으로 하는 방법.

청구항 2

제 1 항에 있어서, 각각의 상기 X의 인간의 위치를 결정하도록 상기 N의 확률들을 이용하는 단계를 더 포함하는 것을 특징으로 하는 방법.

청구항 3

제 2 항에 있어서, 각각의 상기 X의 인간들의 결정된 위치는 상기 비디오 이미지에 대응되는 이미지 평면 내의 위치인 것을 특징으로 하는 방법.

청구항 4

제 2 항에 있어서, 각각의 상기 X의 인간들의 결정된 위치는 실세계에 대응되는 물리적 그라운드 평면(ground plane)에 대한 위치인 것을 특징으로 하는 방법.

청구항 5

제 1 항에 있어서, 상기 비디오 이미지의 전경 픽셀들을 결정하는 단계는 전경 객체들 없는 상기 비디오 이미지의 제1 프레임 및 상기 전경 객체들을 구비하는 상기 비디오 이미지의 제2 프레임의 비교를 포함하는 것을 특징으로 하는 방법.

청구항 6

제 1 항에 있어서, 상기 소정의 형상은 각각의 상기 N의 소정의 위치들에 대해 동일한 것을 특징으로 하는 방법.

청구항 7

제 1 항에 있어서, 상기 N의 소정의 위치들의 적어도 일부에 대한 상기 소정의 형상은 다른 크기를 가지는 것을 특징으로 하는 방법.

청구항 8

제 7 항에 있어서, 각각의 상기 N의 소정의 위치들에 대한 상기 소정의 형상의 크기는 비디오 시스템의 보정에 대응하여 결정되며,

상기 비디오 시스템은 상기 비디오 이미지를 수득하는 데 사용되는 것을 특징으로 하는 방법.

청구항 9

제 8 항에 있어서, 상기 비디오 시스템의 보정은 각각의 상기 N의 소정의 위치들에서 평균 인간 크기에 대응하는 상기 비디오 이미지의 일부의 이미지 크기를 결정하는 단계를 포함하며,

각각의 N의 소정의 위치들에 대한 상기 소정의 형상의 크기는 상기 대응되는 이미지 크기에 대응하여 결정되는 것을 특징으로 하는 방법.

청구항 10

제 1 항에 있어서, 전경 픽셀들(foreground pixels)인 상기 비디오 이미지의 픽셀들의 결정 이전에, 각각의 상기 N의 소정의 위치들에 대하여, 인간이 상기 대응되는 위치에 존재할 때에 상기 비디오 이미지 내에 점유되는 전경 이미지 부분을 추정하여 상기 대응되는 소정의 형상을 결정하는 단계를 더 포함하는 것을 특징으로 하는 방법.

청구항 11

제 10 항에 있어서, 각각의 상기 N의 소정의 위치들에 대한 상기 전경 이미지 부분을 추정하는 단계는 상기 비디오 이미지의 이미지 평면상으로의 실세계 내의 인간의 모델의 투영에 기초하여 계산되는 것을 특징으로 하는 방법.

청구항 12

제 1 항에 있어서, 상기 비디오 이미지는 복수의 이미지 프레임들을 포함하고, 각 이미지 프레임은 상기 N의 소정의 위치들을 갖는 2차원 이미지를 포함하며, 각각의 상기 N의 소정의 위치들은 상기 2차원 이미지 내의 대응되는 x, y 좌표 쌍에 의해 식별되는 것을 특징으로 하는 방법.

청구항 13

제 12 항에 있어서, 각각의 상기 N의 소정의 위치들은 상기 비디오 이미지에 대응되는 이미지 평면에 대하여 상기 N의 소정의 형상들의 대응되는 것과 연관되는 것을 특징으로 하는 방법.

청구항 14

제 1 항에 있어서, 각각의 상기 N의 소정의 위치들에 대하여, 연관 확률을 결정하도록 상기 대응되는 소정의 형상 및 상기 전경 블롭 세트의 리콜 비율(recall ratio)을 계산하는 단계를 더 포함하는 것을 특징으로 하는 방법.

청구항 15

제 14 항에 있어서, 각각의 상기 N의 소정의 위치들에 대하여, 상기 리콜 비율을 계산하는 단계는 (a) 상기 소정의 형상 및 상기 전경 블롭 세트에 의해 점유되는 면적의 중첩을 포함하는 면적 및 (b) 상기 소정의 형상에 의해 점유되는 면적의 비율을 결정하는 단계를 포함하는 것을 특징으로 하는 방법.

청구항 16

제 1 항에 있어서,

상기 N의 확률들을 가지는 확률 맵을 생성하는 단계; 및

상기 확률 맵의 확률들의 국지적인 최대들을 결정하는 단계를 더 포함하는 것을 특징으로 하는 방법.

청구항 17

제 16 항에 있어서,

상기 확률 맵의 국지적인 최대에 대응되는 상기 N의 소정의 위치들의 제1 위치를 선택하는 단계;

상기 제1 위치에 대응되는 제1의 소정의 형상을 수득하는 단계; 및

상기 제1의 소정의 형상 및 상기 전경 블롭 세트에 의해 점유되는 면적의 중첩의 양을 분석하는 단계를 더 포함

하는 것을 특징으로 하는 방법.

청구항 18

제 16 항에 있어서,

상기 확률 맵의 국지적인 최대에 대응되는 상기 N의 소정의 위치들의 제1 위치를 선택하는 단계;

상기 제1 위치에 대응되는 제1의 소정의 형상을 수득하는 단계; 및

(a) 상기 제1의 소정의 형상 및 상기 전경 블롭 세트에 의해 점유되는 면적의 중첩을 포함하는 면적 및 (b) 상기 제1의 소정의 형상의 면적의 제1 비율을 계산하는 단계를 더 포함하는 것을 특징으로 하는 방법.

청구항 19

제 17 항에 있어서,

(a) 상기 제1의 소정의 형상 및 상기 전경 블롭 세트에 의해 점유되는 면적의 중첩을 포함하는 면적 및 (b) 상기 전경 블롭 세트의 면적의 제1 비율을 계산하는 단계를 더 포함하며,

상기 제1 비율은 X의 인간들이 상기 전경 블롭 세트에 의해 나타나는 것을 결정하도록 사용되는 것을 특징으로 하는 방법.

청구항 20

제 1 항에 있어서,

상기 N의 확률들에 기초하여 상기 N의 소정의 형상들의 서브세트를 선택하는 단계; 및

상기 N의 소정의 형상들의 선택된 서브세트에 의해 점유되는 면적과 상기 전경 블롭 세트에 의해 점유되는 면적의 중첩을 분석하는 단계를 더 포함하는 것을 특징으로 하는 방법.

청구항 21

제 16 항에 있어서, 상기 N의 소정의 위치들의 각각의 m(m은 정수이다)의 위치들에 대한 정확도 값 및 리콜 값을 계산하는 단계를 더 포함하며, 각각의 상기 m의 위치들은 상기 확률 맵의 국지적인 최대에 대응되는 것을 특징으로 하는 방법.

청구항 22

제 21 항에 있어서, 1부터 m까지 연속하여 상기 N의 소정의 위치들의 것들을 선택하는 단계를 더 포함하며, (m-1)번째 위치의 선택은 상기 (m-1)번째 위치의 제1의 소정의 거리 내에 있는 상기 N의 소정의 위치들의 후속되는 것의 선택을 제외하는 것을 특징으로 하는 방법.

청구항 23

제 22 항에 있어서, 1부터 m까지 연속하여 상기 N의 소정의 위치들의 것들을 선택하는 단계를 더 포함하며, 상기 N의 소정의 위치들의 다음의 위치의 선택은 상기 비디오 이미지의 하부 에지에 대한 그 근접에 기초하여 위치를 선택하는 단계를 포함하는 것을 특징으로 하는 방법.

청구항 24

비디오 내의 인간 객체들을 탐지하는 방법에 있어서,

실세계 장면의 비디오 이미지의 픽셀들을 전경 픽셀들로 결정하는 단계를 포함하고, 상기 전경 픽셀들의 그룹은 하나 또는 그 이상의 전경 블롭들의 전경 블롭 세트를 구성되며;

상기 비디오 이미지 내의 N(여기서 N은 1보다 큰 정수이다)의 소정의 위치들의 대응되는 것들에서 각각의 N의 소정의 형상들에 대하여, 상기 전경 블롭 세트에 의해 나타나는 X(여기서 X는 전체 숫자이다)의 인간들을 결정하도록 상기 대응되는 소정의 형상을 상기 전경 블롭 세트와 비교하는 단계를 포함하고, 각각의 상기 X의 인간들의 위치는 실세계의 수평 평면 내의 위치로서 결정되며,

상기 X의 인간들의 표시의 결정을 이용하여 군중 밀도(crowd density)가 쓰레시홀드(threshold)를 초과하는 때

에 보고, 경보 및 사건 검출의 적어도 하나를 제공하는 단계를 포함하는 것을 특징으로 하는 방법.

청구항 25

제 24 항에 있어서, 상기 X의 인간들의 위치들의 적어도 일부를 검토하여 군중의 존재를 탐지하는 단계를 더 포함하는 것을 특징으로 하는 방법.

청구항 26

제 24 항에 있어서, 상기 X의 인간들의 Y가 상기 실세계의 수평 평면의 제1 면적 내에 위치하는 것이 결정되는 때에 군중의 존재를 결정하는 단계를 더 포함하는 것을 특징으로 하는 방법.

청구항 27

제 26 항에 있어서, 상기 제1 면적은 상기 실세계 내의 소정의 면적을 갖는 소정의 기하학적 형상을 포함하는 것을 특징으로 하는 방법.

청구항 28

제 26 항에 있어서, 상기 제1 면적은 원에 의해 정의되는 면적을 포함하는 것을 특징으로 하는 방법.

청구항 29

제 26 항에 있어서, 상기 제1 면적 내의 군중 밀도(crowd density)를 결정하는 단계를 더 포함하는 것을 특징으로 하는 방법.

청구항 30

제 29 항에 있어서, 상기 군중 밀도를 쓰레시홀드와 비교하는 단계 및 상기 군중 밀도가 상기 쓰레시홀드를 초과하는 때에 적어도 하나의 보고 및 경보를 전송하는 단계를 더 포함하는 것을 특징으로 하는 방법.

청구항 31

제 24 항에 있어서,

상기 비디오 이미지의 제1 프레임에 대응되는 제1 면적 내의 제1 군중 밀도를 결정하는 단계;

상기 비디오 이미지의 제2 프레임에 대응되는 상기 제1 면적 내의 제2 군중 밀도를 결정하는 단계; 및

상기 제1 군중 밀도 및 상기 제2 군중 밀도에 대응하여 군중 집회(crowd gathering) 사건을 결정하는 단계를 더 포함하는 것을 특징으로 하는 방법.

청구항 32

제 24 항에 있어서,

상기 비디오 이미지의 제1 프레임에 대응하는 제1 면적 내의 제1 군중 밀도를 결정하는 단계;

상기 비디오 이미지의 제2 프레임에 대응하는 상기 제1 면적 내의 제2 군중 밀도를 결정하는 단계; 및

상기 제1 군중 밀도 및 상기 제2 군중 밀도에 대응하여 군중 해산(crowd dispersing) 사건을 결정하는 단계를 더 포함하는 것을 특징으로 하는 방법.

발명의 설명

기 술 분 야

[0001] 본 발명은 비디오 감시(video surveillance) 방법들 및 시스템들과 비디오 검증(video verification) 방법들 및 시스템들과 같은 비디오 감시에 관한 것이다. 인간을 탐지할 수 있는 비디오 감시 시스템들, 장치들 및 방법들이 개시된다. 비디오 감시 시스템들, 장치들 및 방법들은 인간들을 계산할 수 있거나 및/또는 비디오 스트림들 내의 인간 군중 시나리오들을 모니터할 수 있다.

[0002] 본 출원은 2013년 3월 15일에 출원된 미국 특허 출원 제13/838,511호 및 2012년 9월 12일에 출원된 미국 임시

특허 출원 제61/700,033호를 우선권들로 수반하며, 이들 각각의 개시 사항들은 여기에 참조로 포함된다.

배 경 기 술

[0003] 지능형 비디오 감시(IVS) 시스템은 실시간으로 또는 오프라인으로(예를 들면, 미리 기록되고 저장된 비디오를 검토하여) 비디오 피드들 내의 관심의 대상인 사건들을 탐지하는 데 이용될 수 있다. 통상적으로 이러한 작업은 관심의 대상인 타겟들을 탐지하고 추적함에 의해 이루어질 수 있다. 이는 통상적으로 장면이 혼잡하지 않을 때에 잘 수행된다. 그러나, 이와 같은 시스템의 성능은 혼잡한 장면들에서 상당히 떨어질 수 있다. 실제로, 이와 같은 혼잡한 장면들이 자주 발생되며, 이에 따라 군중들 내에서 사람들을 탐지할 수 있는 것이 크게 관심의 대상이 되고 있다. 이와 같은 사람들의 탐지는 군중 밀도, 군중 형성 및 군중 해산과 같은 계산 및 다른 군중 분석들을 위해 사용될 수 있다.

[0004] 이전의 군중 분석 작업은 특정 스포츠 또는 종교 사건들과 같은 일부 특정한 극히 혼잡한 시나리오들을 다루었다. 그러나, 대규모의 군중들이 때때로 형성될 수 있는 보다 보편적인 감시 시나리오들에도 초점을 맞출 필요가 있다. 이들은 거리들, 쇼핑센터들, 공항들, 버스 및 기차역들 등과 같은 공공장소들을 포함한다.

[0005] 최근에, 군중 밀도 추정이나 군중 내의 사람들을 계산하는 문제가 연구 단체들에서 뿐만 아니라 산업계로부터 상당한 주목을 받고 있다. 현재의 접근 방식들은 주로 맵(map) 기반의(간접적인) 접근 방식들 및/또는 탐지 기반의(직접적인) 접근 방식들을 포함한다.

[0006] 맵 기반의 접근 방식은 인간 타겟들의 숫자를 움직임 픽셀들의 양, 전경 블롭(foreground blob) 크기, 전경 에지들, 전경 코너들의 그룹 및 다른 이미지 특징들과 같은 추출된 이미지 특징들로 맵핑(mapping)시키는 시도를 할 수 있다. 상기 맵 기반의 접근 방식은 대체로 다른 형태들의 비디오 시나리오들을 위한 트레이닝을 요구한다. 이러한 조사는 사람들의 계수에 잘 대응되는 신뢰성 있는 특징들을 조사하고, 어떻게 쉐도우들(shadows) 및 카메라 뷰 시점과 같은 일부 특정한 문제들을 처리하는 가에 주로 초점이 맞추어져 있다. 많은 시나리오들 하에서, 상기 맵 기반의 접근 방식은 충분히 훈련된 비디오들로 주어지는 상당히 정확한 인간 계산추정들을 제공할 수 있다. 그러나, 수행이 대체로 장면에 의존적이며, 각 개인의 실제 위치들이 사용되지 않을 수 있다.

[0007] 상기 탐지 기반의 접근 방식은 각 개별 인간 타겟을 확인함에 의해 상기 장면 내의 사람들의 숫자를 셀 수 있다. 이러한 조사는 인간 탐지, 인간 부분들의 탐지 및 탐지와 추적으로 공동 고찰에 초점을 맞추어왔다. 이들 접근 방식들은 적게 혼잡한 시나리오들 내에 보다 정확한 탐지 및 계수를 제공할 수 있다. 각 개인의 위치가 사용 가능하게 만들어질 수 있을 경우, 국지적인 군중 밀도를 계산하는 것이 가능해질 수 있다. 이들 접근 방식들의 중요한 과제들은 보다 높은 계산 비용, 학습에 의존적인 관점 및 상대적으로 큰 인간 이미지 요건이다.

[0008] 여기에 기재되는 실시예들은 현존하는 시스템들의 이들 문제들의 일부를 해결한다.

발명의 내용

해결하려는 과제

[0009] 여기에 기재되는 실시예들은 현존하는 시스템들의 이들 문제들의 일부를 해결한다.

과제의 해결 수단

[0010] 개시된 실시예들은 인간 객체들과 같은 객체들을 탐지하도록 비디오 영상들의 지능형 분석을 위한 방법들, 장치들 및 시스템들을 제공한다.

[0011] 특정 실시예들에 있어서, 비디오 내의 인간 객체들(human objects)을 탐지하는 방법은, 비디오 이미지의 특정 픽셀들이 전경 픽셀들(foreground pixels)인 것을 결정하는 단계를 포함하고, 상기 전경 픽셀들의 그룹은 하나 또는 그 이상의 전경 블롭들의 전경 블롭 세트(foreground blob set)를 구성하며; 상기 비디오 이미지 내의 각각의 N의 위치들에 대하여, 여기서 N은 정수이고, 상기 위치에서 인간의 대응되는 확률을 수득하도록 소정의 형상을 상기 전경 블롭 세트와 비교하는 단계를 포함하며, 이에 따라 상기 N의 위치들에 대응되는 N의 확률들을 수득하고; 상기 N의 확률들을 이용하여, 상기 전경 블롭 세트에 의해 나타나는 X의 인간들을 결정하는 단계를 포함하며, 여기서 X는 전체 숫자(whole number)이다.

[0012] 비디오 내의 인간 객체를 탐지하기 위한 방법은, 실세계 장면의 비디오 이미지의 픽셀들이 전경 픽셀들인 것을

결정하는 단계를 포함할 수 있고, 상기 전경 픽셀들의 그룹은 하나 또는 그 이상의 전경 블롭들의 전경 블롭 세트를 구성하며; 상기 비디오 이미지 내의 각각의 N의 위치들에 대하여, 여기서 N은 정수이고, 상기 전경 블롭 세트에 의해 나타나는 X의 인간들을 결정하도록 소정의 형상을 상기 전경 블롭 세트와 비교하는 단계를 포함할 수 있으며, 여기서 X는 전체 숫자이다.

[0013] 방법들은 각각의 상기 X의 인간들의 위치를 결정하는 단계를 포함할 수 있다. 각각의 상기 X의 인간들의 위치들은 상기 실세계의 물리적 그라운드 평면상의 위치와 같이 상기 실세계의 수평 평면 내의 위치로서 결정될 수 있다.

[0014] 인간 객체들의 탐지는 군중 분석들을 위해서와 다른 사건 탐지들을 위해서 인간들을 계산하는 데 이용될 수 있다.

[0015] 이러한 방법들을 수행하도록 구성될 수 있는 시스템 및 장치들이 개시된다.

[0016] 컴퓨터 판독 가능 소프트웨어는 여기에 설명되는 동작들을 수행하도록 컴퓨터를 구성하는 데 사용될 수 있고 본 발명의 다른 실시예들을 포함할 수 있다.

도면의 간단한 설명

[0017] 예시적인 실시예들은 첨부된 도면들과 함께 기술되는 다음의 상세한 설명으로부터 보다 명확하게 이해될 것이다. 첨부된 도면들은 여기에 설시되는 제한적이지 않은 예시적인 실시예들을 나타낸다.

도 1은 본 발명의 예시적인 실시예에 따른 예시적인 비디오 감시 시스템을 나타낸다.

도 2는 본 발명의 예시적인 실시예에 따른 비디오 감시 시스템으로부터의 비디오 스트림으로부터 예시적인 프레임을 나타낸다.

도 3a는 본 발명의 예시적인 실시예에 따른 타겟 탐지 및 계산을 위한 예시적인 흐름도를 나타낸다.

도 3b는 2차원 비디오 이미지에 대하여 다른 위치에 각기 대응되는 2차원 비디오 이미지를 차지하는 실시예를 나타낸다.

도 3c는 대응되는 인간 모델(320)과 각기 연관되는 (x, y) 식별 좌표(321)의 단일 열을 나타낸다.

도 3d는 인간 확률 맵을 계산하기 위한 예시적인 방법을 나타낸다.

도 3e는 비디오 이미지 내의 인간 모델들의 최적의 숫자를 발견하는 일부로서 확률 맵의 단일 패스를 수행하는 예시적인 방법을 나타낸다.

도 3f는 비디오 이미지 내의 인간 모델들의 최적의 숫자의 발견에 대하여 확률 맵의 복수의 패스들을 수행하는 방법을 나타낸다.

도 4는 3D 실린더 모델 및 이의 대응되는 2D 볼록 헐 모델을 포함하는 범용 인간 모델을 나타낸다.

도 5는 몇몇의 인간 이미지 샘플들을 이용하여 계산될 수 있는 범용 플랫-어스 카메라 모델을 나타낸다.

도 6a, 도 6b 및 도 6c는 예시적인 탐지 결과들을 나타낸다.

도 7a, 도 7b 및 도 7c는 인간 탐지 결과들에 기초하는 인간 군중 밀도에 관한 실시예를 나타낸다.

도 8은 다양한 군중 연관 사건들을 탐지하기 위한 예시적인 구현예들을 나타낸다.

도 9는 혼잡한 지역을 어떻게 정의하고 탐지하는 지의 예시적인 방법을 나타낸다.

도 10은 각 탐지된 인간 타겟에 대한 예시적인 프로세스를 나타낸다.

도 11은 각 군중 영역에 대한 예시적인 프로세스를 나타낸다.

도 12는 군중 "집회" 및 "해산" 사건들을 정의하고 탐지하는 데 사용될 수 있는 방법을 나타낸다.

도 13은 군중 집회 스팟을 정의하는 하나의 실시예를 나타낸다.

도 14a 및 도 14b는 군중 집회 스팟의 예를 나타낸다.

도 15는 상기 군중 집회 스팟들을 탐지하는 예시적인 방법을 나타낸다.

도 16은 상기 군중 집회 스팟들을 업데이트하고 군중 "집회" 및 "해산" 사건들을 탐지하는 예시적인 방법을 나타낸다.

도 17은 복수의 비디오 카메라들을 사용하는 예시적인 구현예를 나타낸다.

발명을 실시하기 위한 구체적인 내용

[0018] 다양한 예시적인 실시예들이 일부 예시적인 실시예들이 도시되는 첨부된 도면들을 참조하여 보다 상세하게 기술될 것이다. 그러나, 본 발명은 많은 다른 형태들로 구현될 수 있으며, 여기에 설시되는 예시적인 실시예들에 한정되는 것으로 간주되지 않아야 한다. 이들 예시적인 실시예들은 단지 실시예들과 많은 구현예들 및 변형들이 여기에 제공되는 세부 사항들을 요구하지 않는 것이 가능한 점을 나타낸다. 본 발명이 선택적인 실시예들의 세부 사항들을 제공하지만, 이러한 선택적인 예들의 나열이 철저하지 않은 점도 강조되어야 할 것이다. 또한, 다양한 실시예들 사이의 세부 사항들의 어떤 일치성은 여기에 기재되는 모든 특징들에 대한 모든 가능한 변화를 열거하는 것이 거의 불가능하기 때문에 이러한 세부 사항들을 요구하는 것으로 해석되지 않아야 한다. 특허청구범위의 표현은 본 발명의 요구 사항들의 결정에 참조되어야 한다. 도면들에 있어서, 층들과 영역들의 크기들 및 상대적인 크기들은 명료성을 위해 과장될 수 있다. 도면들에 걸쳐 동일한 참조 부호들은 동일한 요소들을 나타낸다.

[0019] 비록 제1, 제2, 제3 등의 용어들이 여기서 다양한 요소들을 기술하는 데 사용될 수 있지만, 이들 요소들이 이들 용어들에 의해 한정되지 않아야 하는 점이 이해될 것이다. 이들 용어들은 하나의 요소를 다른 하나로부터 구별하는 데 사용된다. 따라서, 다음에 논의하는 제1 요소는 본 발명의 설시로부터 벗어나지 않고 제2의 요소로 호칭될 수 있다. 여기에 사용되는 바에 있어서, "및/또는"이라는 용어는 연관되고 열거되는 항목들의 하나 또는 그 이상의 임의의 및 모든 결합들을 포함한다.

[0020] 요소가 다른 요소에 "연결되는" 또는 "접속되는"으로 언급되는 때에, 이는 또 다른 요소에 직접 연결되거나 접속될 수 있거나, 개재되는 요소들이 존재할 수 있는 점이 이해될 것이다. 대조적으로, 요소가 다른 요소에 "직접 연결되는" 또는 "직접 접속되는"으로 언급되는 때, 개재되는 요소들이 존재하지 않는다. 요소들 사이의 관계를 기술하는 데 사용되는 다른 단어들은 동일한 방식(예를 들면, "사이에" 대 "직접 사이에", "인접하는" 대 "직접 인접하는" 등)으로 해석되어야 한다.

[0021] 여기에 사용되는 용어들은 특정한 예시적인 실시예들을 설시하기 위한 목적만이며 본 발명을 제한하는 것으로 의도되지는 않는다. 여기에 사용되는 바에 있어서, 단수 형태들인 "일" "하나" 및 "상기"는 본문에서 명백하게 다르게 나타내지 않는 한 복수의 형태들로 포함하는 것으로 의도된다. "포함하다" 및/또는 "포함하는"이라는 용어들은 본 명세서에서 사용되는 때에 언급된 특징들, 정수들, 단계들, 동작들, 요소들 및/또는 성분들이 존재를 명시하지만, 하나 또는 그 이상의 다른 특징들, 정수들, 단계들, 동작들, 요소들, 성분들 및/또는 이의 집합들의 존재나 추가를 불가능하게 하는 것은 아닌 점이 더 이해될 것이다.

[0022] 다르게 정의되지 않는 한, 여기서 사용되는 모든 용어들(기술 및 과학 용어들을 포함하여)은 본 발명이 속하는 해당 기술 분야에서 통상의 지식을 가진 자에 의해 통상적으로 이해되는 의미를 가진다. 또한, 공통적으로 사용되는 사전들에서 정의되는 경우들과 같은 용어들은 관련 기술의 분야에서의 이들의 의미와 일치되는 의미를 가지는 것으로 해석되어야 하며, 여기서 명백하게 다르게 정의하지 않는 한 이상적이거나 지나치게 형식적인 의미로 해석되지는 않는 점이 이해될 것이다.

[0023] 정의들

[0024] 본 발명을 설명함에 있어서, 다음 정의들은 전체적으로(상술한 바들을 포함하여) 적용될 수 있다.

[0025] "비디오(video)"는 아날로그 및/또는 디지털 형태로 나타나는 활동사진들을 언급할 수 있다. 비디오의 예들은, 텔레비전; 영화; 비디오 카메라 또는 다른 관찰자로부터의 연속 영상; 라이브 피드(live feed)로부터의 연속 영상; 컴퓨터 생성 연속 영상; 컴퓨터 그래픽 엔진으로부터의 연속 영상; 컴퓨터 판독 가능 매체, 디지털 비디오 디스크(DVD) 또는 고화질 디스크(HDD)와 같은 저장 장치로부터의 연속 영상; IEEE 1394 기반의 인터페이스로부터의 연속 영상; 비디오 디지타이저(video digitizer)로부터의 연속 영상; 또는 네트워크로부터의 연속 영상을 포함할 수 있다.

[0026] "비디오 시퀀스(video sequence)"는 비디오의 일부 또는 전부를 언급할 수 있다.

[0027] "비디오 카메라(video camera)"는 영상 녹화를 위한 장치를 언급할 수 있다. 비디오 카메라의 예들은 다음의 하나 또는 그 이상을 포함할 수 있다. 비디오 이미저 및 렌즈 장치; 비디오 카메라; 디지털 비디오 카메라; 컬러 카메라; 모노크롬(monochrome) 카메라; 카메라; 캠코더; PC 카메라; 웹 캠; 적외선(IR) 비디오 카메라; 저조도 비디오 카메라; 열화상 카메라; 폐쇄 회로 텔레비전(CCTV) 카메라; PZT(pan-tilt-zoom) 카메라; 그리고 비디오 감지 장치. 비디오 카메라는 관심의 대상이 되는 지역의 감시를 수행하기 위해 위치할 수 있다.

[0028] "영상 처리(video processing)"는, 예를 들면, 압축, 편집, 감시 및/또는 검증을 포함하는 영상의 임의의 조작 및/또는 분석을 언급할 수 있다.

[0029] "프레임(frame)"은 영상 내의 특정한 이미지 또는 다른 개별 단위를 언급할 수 있다.

[0030] "컴퓨터(computer)"는 구조화된 입력을 수용할 수 있고, 지정된 규칙들에 따라 상기 구조화된 입력을 처리할 수 있으며, 출력으로서 상기 처리의 결과들을 생성할 수 있는 하나 또는 그 이상의 장치들 및/또는 하나 또는 그 이상의 시스템들을 언급할 수 있다. 컴퓨터의 예들은, 컴퓨터; 거치용 및/또는 휴대용 컴퓨터; 병렬로 및/또는 병렬이 아니게 동작할 수 있는 단일 프로세서, 다중 프로세서들 혹은 다중 코어 프로세서들을 갖는 컴퓨터; 범용 컴퓨터; 슈퍼컴퓨터; 중앙 컴퓨터(mainframe); 슈퍼 미니컴퓨터; 미니컴퓨터; 워크스테이션; 마이크로컴퓨터; 서버; 클라이언트(client); 양방향 텔레비전; 웹 기기(web appliance); 인터넷 접근이 가능한 통신 장치; 컴퓨터와 양방향 텔레비전의 혼성 결합; 휴대용 컴퓨터; 태블릿 개인용 컴퓨터(PC); 개인용 정보 단말기(PDA); 휴대용 전화기; 예를 들면, 디지털 신호 처리기(DSP), 필드-프로그래머블 게이트 어레이(FPGA), 응용 주문형 집적 회로(ASIC), 주문형 응용 명령-세트 처리기(ASIP), 칩, 칩들 또는 칩 세트와 같은 컴퓨터 및/또는 소프트웨어를 실행시키는 특정 용도 지향 하드웨어; 시스템 온 칩(SoC) 또는 멀티프로세서 시스템 온 칩(MPSoC); 광 컴퓨터; 양자 컴퓨터; 바이오컴퓨터; 그리고 데이터를 수용할 수 있고, 하나 또는 그 이상의 저장된 소프트웨어 프로그램들에 따라 데이터를 처리할 수 있으며, 결과들을 발생시킬 수 있고, 통상적으로 입력, 출력, 저장, 연산, 로직 및 제어 유닛들을 구비할 수 있는 장치를 포함할 수 있다.

[0031] "소프트웨어(software)"는 컴퓨터를 동작시키기 위해 규정된 규칙들을 언급할 수 있다. 소프트웨어의 예들은, 소프트웨어; 코드 세그먼트들(code segments); 명령들; 애플릿들(applets); 미리 컴파일된(pre-compiled) 코드; 컴파일된 코드; 해석된 코드; 컴퓨터 프로그램들; 그리고 프로그램된 로직을 포함할 수 있다.

[0032] "컴퓨터 판독 가능 매체(computer-readable medium)"는 컴퓨터에 의해 접근 가능한 데이터를 저장하기 위해 사용되는 임의의 저장 장치를 언급할 수 있다. 컴퓨터 판독 가능 매체의 예들은, 자기 하드 디스크; 플로피 디스크; CD-ROM 및 DVD와 같은 광 디스크; 자기 테이프; 플래시 이동식 메모리; 메모리 칩; 및/또는 내부에 기계 판독 가능 명령들을 저장하는 다른 유형들의 매체들을 포함할 수 있다.

[0033] "컴퓨터 시스템(computer system)"은 하나 또는 그 이상의 컴퓨터들을 갖는 시스템을 언급할 수 있으며, 여기서 각 컴퓨터는 상기 컴퓨터를 동작시키기 위해 소프트웨어를 담는 컴퓨터 판독 가능 매체를 포함할 수 있다. 컴퓨터 시스템의 예들은, 네트워크에 의해 연계된 컴퓨터 시스템들을 통해 정보를 처리하기 위한 분산된 컴퓨터 시스템; 컴퓨터 시스템들 사이에서 정보를 전송 및/또는 수신하기 위해 네트워크를 통해 함께 연결된 둘 또는 그 이상의 컴퓨터 시스템들; 그리고 데이터를 수용할 수 있고, 하나 또는 그 이상의 저장된 소프트웨어 프로그램들에 따라 데이터를 처리할 수 있으며, 결과들을 발생시킬 수 있고, 통상적으로 입력, 출력, 저장, 연산 및 제어 유닛들을 구비할 수 있는 하나 또는 그 이상의 장치들 및/또는 하나 또는 그 이상의 시스템들을 포함할 수 있다.

[0034] "네트워크(network)"는 통신 설비들에 의해 연결될 수 있는 많은 컴퓨터들 및 관련 장치들을 언급할 수 있다. 네트워크는 케이블들과 같은 영구적인 연결들이나 전화 또는 다른 통신 링크들을 통해 구현되는 경우들과 같은 일시적인 연결들을 수반할 수 있다. 네트워크는 고정 배선 연결들(예를 들면, 동축 케이블, 트위스트 페어(twisted pair), 광섬유, 도파관들 등) 및/또는 무선 연결들(예를 들면, 무선 주파수 파형들, 자유 공간 광 파형들, 음향 파형들 등)을 더 포함할 수 있다. 네트워크의 예들은, 인터넷(Internet)과 같은 인터넷; 인트라넷; 근거리 통신망(LAN); 광역통신망(WAN); 그리고 인터넷 및 인트라넷과 같은 네트워크들의 결합을 포함할 수 있다. 예시적인 네트워크들은 인터넷 프로토콜(IP), 비동기 전송 방식(ATM) 및/또는 동기식 광 네트워크(SONET), 사용자 데이터그램 프로토콜(UDP), IEEE 802.x 등과 같은 다수의 프로토콜들(protocols)의 임의의 것으로 동작할 수 있다.

[0035] 일부 실시예들에 있어서, 군중 밀도(crowd density) 추정 방법, 시스템 및 장치는 현존하는 비디오 내용 분석 방법들, 시스템들 및 장치들에 기초할 수 있다. 기본적인 추정 정확도 요구이외에도 상기 접근 방식은 다음의

하나 또는 그 이상을 포함할 수 있다.

[0036] ● 카메라 뷰(camera view) 독립은 실시예들이 카메라 위치, 화각(view angle), 타겟 상의 픽셀들의 수 등의 변화들에 관계없이 넓은 범위의 응용 시나리오들 상에서 작동되게 할 수 있다.

[0037] ● 실시간으로 수행될 수 있는 상대적으로 낮은 연산 비용. 상기 실시예들은 임베디드 시스템(embedded system) 상에 구현될 수 있다.

[0038] ● 복잡한 초기 설정 및 트레이닝이 감소될 수 있거나 및/또는 제거될 수 있으므로 보다 편리하며 낮은 소유 비용을 가능하게 한다.

[0039] 여기에 개시되는 일부 예들은 탐지 기반의 접근 방식을 포함하며, 트레이닝이 요구되지 않을 수 있다. 상기 예들은 이미 기본 탐지와 추적 작업들을 수행하고 신뢰성 있는 전경 마스크(foreground mask)를 제공하는 일반적인 IVS 시스템으로 구현될 수 있다. 볼록한 영역(convex region) 인간 이미지 모델은 모든 이미지 픽셀에 대해 계산될 수 있으며, 이는 각 전경 면적 내의 인간 타겟들의 숫자를 추정하는 데 이용될 수 있다. 카메라 보정 데이터는 맵핑(mapping)을 이미지 평면으로부터 물리적 세계의 그라운드 평면까지 제공할 수 있으며, 이는 카메라 뷰 내의 영역들에서 실제 군중 밀도 측정들을 제공하는 데 이용될 수 있다. 상기 실제 군중 밀도 측정(들)을 이용하여, 예를 들면, "군중 핫 스팟(crowd hot spot)", "군중 집회(crowd gathering)", "군중 해산(crowd dispersing)" 등의 관심의 대상인 다른 사건들이 검출될 수 있다.

[0040] 도 1은 본 발명의 예시적인 실시예들에 따른 비디오 감시 시스템(101)을 나타낸다. 상기 비디오 감시 시스템은 비디오 스트림들 내의 인간 군중 활동들을 탐지하고 모니터하도록 구성될 수 있다. 상기 비디오 감시 시스템(101)은 군중 밀도 분석들을 위한 사용과 같이 인간 탐지가 관심의 대상인 다양한 응용들에 사용될 수 있다. 예를 들면, 실시예들은 수상한 사람 집회 탐지, 보행자 통행량 통계 수집, 비정상적인 군중 형성 및/또는 해산 등에 대해 이용될 수 있다. 상기 비디오 감시 시스템(101)은 비디오 소스(102)(예를 들면, 하드 드라이브와 같은 저장된 비디오를 갖는 비디오 카메라 또는 메모리), 변화 탐지(change detection) 모듈(103), 움직임 탐지(motion detection) 모듈(104), 전경 블롭 탐지(foreground blob detection) 모듈(105), 인간 탐지 모듈(106), 타겟 추적(target tracking) 모듈(107) 및 사건 탐지(event detection) 모듈(108)을 포함할 수 있다. 이러한 예에서, 상기 비디오 소스(예를 들면, 비디오 카메라)는 고정된다. 그러나, 통상의 지식을 가진 자라면 본 발명이 이동하는 비디오 소스들에도 작용되는 점을 이해할 것이다. 이러한 예에서, 상기 비디오 소스는 단일 비디오 스트림을 제공한다. 그러나, 본 발명에는 다중 비디오 스트림들의 사용 및 처리도 고려된다.

[0041] 상기 비디오 감시 시스템은 통상적인 고정된 플랫폼 IVS 시스템으로 구현될 수 있다. 여기에 설시되는 실시예들을 구현하는 데 이용될 수 있는 IVS 시스템의 예시적인 세부 사항들을 위한 예로서, Venetianer 등에게 허여된 미국 특허 제7,868,912호 및 Lipton 등에게 허여된 미국 특허 제7,932,923호를 참조하기 바라며, 이들의 개시 사항들은 여기에 참조로 포함된다. 미국 특허 제7,868,912호 및 미국 특허 제7,932,923호는 또한 사건 탐지와 같이 상기 비디오로부터 정보를 수득하도록 비디오 프리미티브(video primitive)(또는 메타데이터) 발생 및 다운스트림(downstream) 처리(실시간 처리나 이후의 처리가 될 수 있다)의 예시적인 세부 사항들을 위한 참조로 포함되어 상기 생성된 비디오 프리미티브들을 이용하여 사건 탐지와 같이 상기 비디오로부터 정보를 수득하며, 이는 여기에 개시되는 실시예들과 함께 사용될 수 있다. 단독으로 또는 다른 모듈들/구성 요소들과 결합되는 각 모듈(103-108)뿐만 아니라 이들의 개별적인 구성 요소들은 전용의 하드웨어(회로부), 소프트웨어 및/또는 펌웨어에 의해 구현될 수 있다. 예를 들면, 소프트웨어로 프로그램된 범용 컴퓨터가 상기 모든 모듈들을 실행시킬 수 있다. 이와 같이, 여기에 기재되는 동작들을 수행하도록 컴퓨터를 구성하는 데 이용될 수 있는 소프트웨어를 포함하는 컴퓨터 판독 가능 매체는 본 발명의 다른 실시예들을 포함한다. 다른 예로서, 여기에 기술되는 시스템들, 장치들 및 방법들을 구현하기 위하여, 다음의 하나 또는 그 이상과 같은 다양한 컴퓨팅 및 광학 구성 요소들이 사용될 수 있다. 범용 컴퓨터; 슈퍼컴퓨터; 중앙 컴퓨터; 슈퍼 미니컴퓨터; 미니컴퓨터; 워크스테이션; 마이크로 컴퓨터; 서버; 양방향 텔레비전; 컴퓨터와 양방향 텔레비전의 혼성 결합; 스마트 폰; 태블릿; 그리고 컴퓨터 및/또는 소프트웨어를 대리 실행하는 응용 주문형 하드웨어. 이들은 하나 또는 그 이상의 프로세서들, 하나 또는 그 이상의 필드 프로그래머블 게이트 어레이(FPGA)들, 컴퓨터 메모리, 예를 들면, 컴퓨터에 의해 접근 가능한 데이터를 저장하기 위한 임의의 저장 장치(예를 들면, 프로세서는 카메라 장치로부터 수신되는 데이터 상의 다양한 알고리즘들을 수행할 수 있으며, 컴퓨터 메모리는 이후에 다양한 픽셀들에 대한 정보를 저장할 수 있고, 블롭 탐지, 표적 탐지 및 사건 탐지의 결과들을 저장할 수 있다)와 같은 컴퓨터 판독 가능 매체를 포함할 수 있다. 컴퓨터 판독 가능 매체의 예들은, 자기 하드 디스크; 플로피 디스크; CD-ROM 및 DVD와 같은 광 디스크; 자기 테이프; 메모리 칩; 고체 상태 저장 장치(SSD); 그리고 이메일을 전송하고 수신하거나 네트워크에

의 액세싱에 이용되는 경우들과 같이 컴퓨터 판독 가능 전자 데이터를 운송하는 데 이용되는 반송파(carrier wave)를 포함한다. 실재하는 컴퓨터 판독 가능 매체는 앞서 열거한 바와 같은 물리적으로 만질 수 있는 컴퓨터 판독 가능 매체를 포함한다. 또한, 소프트웨어는 여기에 설명되는 방법들을 구현하기 위해 상기 컴퓨팅 및/또는 광학 구성 요소들과 결합되어 사용될 수 있다. 소프트웨어는 컴퓨터를 동작시키는 규칙들 및/또는 알고리즘들을 포함할 수 있고, 예를 들면, 코드 세그먼트들, 명령들, 컴퓨터 프로그램들 및 프로그램된 로직을 포함할 수 있다. 상기 비디오 소스(102) 및 모듈들(103-108)은 단일 시스템 내에 있을 수 있거나 분산될 수 있다. 예를 들면, 비디오 소스(102)는 모니터되는 장소에 비디오 카메라를 포함할 수 있다. 비디오 소스(102)는 모듈들(103-107)이 위치하는 모니터링 장소(예를 들면, 모니터되는 장소로부터 떨어진 별도의 제2의 장소)에 비디오 스트림을 제공한다. 사건 탐지 모듈(108)은 상기 모니터링 장소 및 상기 제2의 장소로부터 분리된 제3의 장소(예를 들면, 중앙 정차장)에 제공될 수 있다. 상기 다양한 모듈들, 컴퓨터들, 카메라들 및 여기에 기재되는 다른 영상 장비는 케이블들과 같은 영구적인 연결들이나 전화 또는 다른 통신 링크들을 통해 구현되는 경우들과 같은 일시적인 연결들을 수반할 수 있고, 무선 통신 링크들로 포함할 수 있는 네트워크상에서 연결될 수 있다. 네트워크의 예들은, 인터넷(Internet)과 같은 인터넷; 인트라넷; 근거리 통신망(LAN); 광역 통신망(WAN); 및 인터넷과 인트라넷과 같은 네트워크들의 결합을 포함한다. 전술한 다양한 하드웨어와 소프트웨어 예들은 또한 여기에 포함되는 특허 문헌들에 상세하게 기재되어 있다.

[0042] 변화 픽셀들은 미리 얻어진 배경 이미지와 다른 비디오 소스(102)에 의해 제공되는 상기 비디오 이미지의 픽셀들로서 상기 변화 탐지 모듈(103)에 의해 탐지될 수 있다. 상기 배경 이미지는 동적일 수 있다. 상기 동적 배경 이미지 모델은 유입되는 비디오 프레임들로부터 계속적으로 구성되고 업데이트될 수 있다. 따라서, 상기 비디오 이미지를 변경하는 조명, 날씨 등의 변화들이 상기 배경 이미지 내에 고려될 수 있다. 104에서, 프레임 차이(frame differencing)가 이동하는 픽셀들을 탐지하는 데 이용될 수 있다. 105에서, 모듈(103)로부터의 변화 픽셀들 및 모듈(104)로부터의 이동하는 픽셀들의 하나 또는 모두가 전경 블롭들 내로 공간적으로 그룹화되는 전경 픽셀들을 결정하기 위해 고려된다. 상기 비디오 이미지는 그 개시 사항이 여기에 참조로 포함되는 Zhang 등에게 허여되고 2010년 11월 2일에 공개된 미국 특허 제7,825,954호에 기재된 바와 같이 전경, 전경 블롭들 및 관심의 대상인 전경 블롭들(인간 전경 블롭들과 같은)을 추출하도록 현재의 비디오 내용 분석 시스템들 및 방법들에 의해 처리될 수 있다. 깊이 센서(depth sensor) 정보는 실세계의 높이 또는 잠재적인 인간으로 탐지되는 각 객체의 크기를 추정하는 데 선택적으로 이용될 수 있으며, 그 결과 잠재적인 인간 타겟들(관심의 대상이 아닌 블롭들과 대조되는 것으로서)에 대응되는 상기 블롭들이 보다 정확하게 확인될 수 있다. 깊이 센서 정보는 쉐도우들(shadows), 반영들(specularities), 관심의 대상인 영역의 외부로서 탐지되는 객체들, 멀리 떨어진 객체들(예를 들면, 정확한 분석들을 허용하기에 충분히 가깝지 않을 수 있는), 또는 상기 비디오 이미지의 잘못된 분석의 위험을 증가시킬 수 있는 상기 비디오 이미지의 다른 요소들을 제거하도록 선택적으로 사용될 수 있다. 깊이 정보의 사용의 예시적인 세부 사항들은 그 개시 사항들이 여기에 참조로 포함되는 Zhang 등에 의한 미국 특허 출원 제13/744,254호에서 찾아볼 수 있다. 상기 블롭들은 타겟 추적 모듈(107) 내의 시공간(spatio-temporal) 타겟들을 형성하도록 시간에 따라 추적되며, 최종적으로 사건 탐지 모듈(108)이 상기 타겟 탐지 및 추적 프로세스의 출력을 이용하여 사용자에 의해 정의되는 관심의 대상인 사건을 탐지한다. 블롭들 내로의 전경 픽셀들의 간단한 공간적인 그룹화(grouping) 대신에 또는 추가적으로, 인간 탐지 모듈(106)은 혼잡한 시나리오들 내의 인간들 사건을 탐지하도록 보정 정보 및 볼록한 영역 형상의 인간 모델을 이용한다. 일부 예들에서, 상기 장면 내의 상기 인간 객체들을 탐지하기 위해 트레이닝이 필요하지 않거나 최소한의 트레이닝이 사전에 요구된다. 또한, 상기 사건 탐지 모듈(108)에서, 인간 탐지 모듈(106)의 결과로 되는 인간 탐지를 이용할 수 있는 일부 새로운 사건 탐지 접근 방식들이 구현될 수 있다.

[0043] 도 2는 실외 광장들, 거리들, 관광 명소들, 기차역들, 쇼핑몰들, 지하철역들 등을 포함하는 상기 IVS 시스템(101)을 위한 일부 통상적인 응용 시나리오들에 대응되는 비디오 이미지들을 나타낸다. 알 수 있는 바와 같이, 녹화되는 장면에 대한 상기 카메라의 위치에 따라, 사람들이 차지하는 비디오 이미지들의 상대적인 크기와 형상이 다르다.

[0044] 도 3a는 상기 비디오 감시 시스템(101)의 보다 예시적인 세부 사항들을 제공하는 블록도를 나타낸다. 전경 블롭 탐지 모듈(105)은 도 1의 경우와 동일할 수 있다. 모듈들(301, 302, 303, 304, 305, 306)은 도 1의 인간 탐지 모듈(106)의 요소들일 수 있다. 인체 픽셀 탐지 모듈(301)은 변화 탐지 모듈(103)로부터의 변화 픽셀 결과들에 기초하여 인체 픽셀들을 탐지한다. 이들 픽셀들은 상기 배경 이미지 모델과 상당히 다르거나(예를 들면, 각 쓰레시홀드(threshold)를 초과하는 휘도 차이 및/또는 색상 차이), 높은 확신의 전경 에지 픽셀들 사이에 위치한다. 이들은 상기 이미지 내의 가장 적당한 인체 픽셀들이 되는 것으로 간주된다. 예를 들면, 탐지된 인체 픽셀들의 예로서 도 6a의 301a를 참조하기 바란다. 다른 변화 픽셀들은 이들이 쉐도우들이나 반영들을 가장 잘 나타

낼 것이기 때문에 다른 인간 탐지 처리로부터 제외될 수 있다. 인간 경계 화소 탐지(human boundary pixel detection) 모듈(302)은 상기 전경 블롭들의 경계가 현재의 비디오 프레임의 이미지 에지들과 정렬되는 인간 경계 화소들을 탐지한다. 예를 들면, 탐지된 인간 경계 픽셀들의 예로서 도 6a의 302a를 참조하기 바란다. 인간 탐지를 수행할 때, 인체가 탐지되었는지의 판단을 보조하도록 다른 분석들이 구현될(상술한 바에 추가적으로 또는 대체하여) 수 있다. 예를 들면, 각 잠재적인 인간 블롭이 경계 전경 에지 픽셀들의 특정 숫자를 포함해야 하는 것이 요구될 수 있다. 다른 예로서, 다른 처리는 인간 보다는 객체(운반체와 같이)와 관련될 가능성이 있는 경우에 블롭(들)을 인식할 수 있으며, 다른 인간 탐지 처리로부터 이와 같은 블롭(들)을 제외할 수 있다. 잠재적인 인간으로 간주되지 않는 다른 전경 블롭들은 상기 전경 블롭 세트로부터 제외될 수 있다. 선택적으로는, 임의의 탐지된 블롭이 상기 전경 블롭 세트의 일부일 수 있다.

[0045] 범용 인간 모델 모듈(303)은 범용 인간 3차원(3D) 및 2차원(2D) 모델을 제공한다. 예를 들면, 상기 범용 인간 모델 모듈(303)은 실세계 내의 3D 인간 모델을 상기 비디오 이미지의 2D 이미지 평면으로 맵핑(mapping)하거나 투영하여 3D 인간 모델을 2D 인간 모델로 변환시킬 수 있다. 이미지 평면(330) 상의 대응되는 2D 인간 모델(303b)로 맵핑되는 예시적인 3D 모델(303a)을 나타낸다. 상기 3D 인간 모델(303a) 실린더들의 그룹(예를 들면, 다리들을 위한 하나의 실린더, 동체를 위한 하나의 실린더 및 머리를 위한 하나의 실린더)과 같은 단순한 3D 형상들의 세트가 될 수 있다. 동일한 3D 인간 모델(303a)(예를 들면, 상기 실린더 모델)이 다양한 비디오 카메라 위치들과 함께 사용될 수 있으므로, 그라운드(실세계의 그라운드 평면)에 대한 상기 비디오 카메라의 다른 각도들이 상기 비디오 카메라의 이미지 평면 내에 다른 형상의 2D 인간 모델(303b)을 수득하도록 이용될 수 있다. 예를 들면, 3D 실린더 인간 모델을 예로서 취하면, 특정한 위치의 탑-다운 뷰를 제공하는 카메라 각도는 상기 2D 이미지 평면 내에 원형으로 맵핑될 수 있는 반면, 동일한 방향의 사면 뷰(oblique view)를 갖는 카메라 각도는 3D 실린더 인간 모델을 연장된 형태를 갖는 다른 형상으로 맵핑시킬 수 있다. 도 17에 도시한 예에서, 카메라(1702)는 카메라(1704)에 비하여 3D 인간 모델(303a)의 탑-다운 뷰를 보다 더 가지므로, 카메라(1702)와 비교하여 3D 인간 모델(303a)의 사이드 뷰를 더 가질 수 있다. 3D 인간 모델(303a)로부터 상기 카메라들(1702, 1704)의 거리들일 동일할 경우, 상기 카메라(1702)의 이미지 평면으로 맵핑되는 대응되는 2D 인간 모델은 상기 카메라(1704)의 이미지 평면으로 맵핑되는 2D 인간 모델 보다 콤팩트하게(예를 들면, 보다 짧게) 된다. 상기 2D 인간 모델은 상기 2D 이미지 평면으로의 상기 3D 인간 모델의 투영의 외부 에지들의 보간 포인트들(interpolating points)에 의해 얻어질 수 있는 볼록한 형상을 가질 수 있다.

[0046] 도 4는 3D 실린더 모델(303a) 및 2D 이미지 평면(330)으로 맵핑되는 이의 대응되는 2D 볼록 헐 모델(convex hull model)(303b)을 포함하는 범용 인간 모델을 예시한다. 상기 3D 인간 모델(303a)은 다리 실린더, 몸통 실린더 및 머리 실린더로 구성된다. 각 실린더의 길이와 반경은 통상적인 보통 인간의 통상적인 치수들을 나타내는 물리적인 통계 데이터에 대응될 수 있다. 도 4에 도시한 바와 같이, 이들 세 실린더들은 머리 평면, 어깨 평면, 엉덩이 평면 및 발 평면의 네 키(key) 평면들을 가진다. 특정한 위치에서 대응되는 2D 인간 모델을 수득하기 위하여, 상기 네 키 평면들의 주위를 따라 균일하게 샘플링할 수 있고, 상기 2D 이미지 공간 내의 특정한 위치에 대한 적절한 크기와 배향을 결정하도록 각 3D 샘플 포인트를 상기 카메라 보정 변수들을 이용하여 상기 2D 이미지 평면상으로 투영시킬 수 있다. 이들 대응되는 이미지 샘플 포인트들은 이후에 상기 2D 영상 인간 모델로서 이용될 수 있는 볼록 형성(convex formation) 방법을 통해 상기 이미지 상에 볼록 헐을 형성하도록 사용될 수 있다.

[0047] 도 5는 몇몇의 인간 이미지 샘플들을 이용하여 보정될 수 있는 범용 플랫-어스(flat-earth) 카메라 모델을 예시한다. 상기 카메라 모델은, 그라운드에 대한 카메라 높이, 이의 틸트-업(tilt-up) 각도 및 상기 카메라의 초점 길이의 세 변수들만을 포함할 수 있다. 이들 변수들은, 각각의 개시 사항들이 여기에 참조로 포함되는 Z. Zhang, P. L. Venetianer 및 A. J. Lipton의 "A Robust Human Detection and Tracking System Using a Human-Model-Based Camera Calibration"(The 8th International Workshop on Visual Surveillance, 2008) 그리고 2010년 9월 21일에 공개된 Zhang 등에게 허여된 미국 특허 제7,801,330호에 기재된 바와 같이, 상기 비디오 프레임들로부터 셋 또는 그 이상의 인간 샘플들을 이용하여 추정될 수 있다.

[0048] 선택적인 또는 추가적인 예에 있어서, 상기 범용 인간 모델 모듈(303)은 상기 비디오 이미지를 취하는 상기 비디오 카메라의 카메라 각도에 대응하여 변경될(예를 들면, 늘어나거나, 수축되거나, 상기 2D 이미지 평면의 수직 축에 대해 기울어지는 등) 수 있는 소정의 2D 모델을 가질 수 있다. 몇몇 범용 인간 모델들은 범용 인간 모델 모듈(303)에 의해 제공될 수 있다. 인간 모델들은 또한 통상적인 액세서리들을 위한 모델링을 포함할 수 있다. 예를 들면, 상기 시스템을 실외에서 사용할 때, 제1의 인간 모델은 따뜻한 날씨에 대해 사용될 수 있고, 제2의 보다 큰 인간 모델은 차가운 날씨에 대해 사용될 수 있으며(코트를 입는 것이 예상되고 상기 인간 모델의

부분으로서 고려되는 경우), 제3의 인간 모델은 비가 오는 날씨(우산이 사용되는 것이 예상되고 상기 인간 모델의 일부로서 고려되는 경우)에 대해 사용될 수 있다.

[0049] 범용 인간 모델 모듈(303)은 또한 상기 이미지 평면 내부의 대응되는 위치들에서 상기 2D 인간 모델의 다양한 크기들의 추정을 제공한다. 상기 이미지 공간은 비디오 소스(102)에 의해 제공되는 비디오의 프레임 내의 이미지의 2차원 공간에 대응될 수 있다. 이미지 공간은 픽셀 증가들에서 측정될 있으므로, 상기 이미지 공간 내의 위치들이 픽셀 좌표들에 의해 확인된다. 비디오 카메라는 3차원 실세계의 2차원 이미지를 포함하는 비디오 이미지를 취할 수 있다. 인간이 실세계의 특정 위치에 존재할 때, 상기 인간은 상기 2차원 비디오 이미지 내부의 특정 위치에서 전경의 특정 양을 차치하는 것으로 예상될 수 있다. 상기 인간이 상기 비디오 카메라로부터 멀리 떨어질 경우, 상기 인간의 이미지 크기는 상기 비디오 카메라에 가까운 인간의 이미지 크기와 비교하여 상대적으로 작은 것으로 예상될 수 있다. 상기 2차원 비디오 이미지 공간 내부의 복수의 위치들의 각각을 위하여, 범용 인간 모델 모듈(303)은 상기 2차원 이미지 공간 내의 위치에 대응되는 크기를 갖는 인간 모델을 제공할 수 있다. 각 위치를 위하여, 상기 2D 인간 모델은 상기 2차원 비디오 이미지의 이미지 공간 내의 해당 위치에 대응하여 치수들 및/또는 크기를 가질 수 있다. 이들 인간 모델들의 배향도 상기 2차원 이미지 공간 내의 위치에 대응할 수 있다. 예를 들면, 일부 카메라 렌즈들(예를 들면, 광각 렌즈들)은 상기 비디오 이미지 프레임의 일 측부에서의 제1의 방향 및 상기 비디오 이미지 프레임의 다른 측부에서 제2의 방향으로 다른 상기 실세계의 수직 방향을 나타낼 수 있다. 상기 2D 인간 모델들은 실세계 수직 방향의 다른 표시들에 대응하여 상기 비디오 이미지 프레임(및 다른 위치들)의 다른 측부들에서 다른 배향들을 가질 수 있다.

[0050] 상기 2D 비디오 이미지 공간 내의 각각의 상기 복수의 인간 모델들의 위치들은 상기 2D 비디오 이미지 공간 내의 식별 좌표와 연관될 수 있다. 상기 식별 좌표는 상기 2D 비디오 이미지 공간을 갖는 비디오의 픽셀 위치들에 대응될 수 있다. 예를 들면, 픽셀 어레이의 10번째 열(row), 22번째 행(column)에 대응되는 위치는 (10,22)의 식별 좌표에 대응될 수 있다. 상기 2D 비디오 이미지 공간 내의 복수의 위치들의 각각을 위하여, 상기 범용 인간 모델 모듈(303)은 상기 인간 모델의 특정한 포인트를 연관된 식별 좌표로 맵핑시킬 수 있다. 예를 들면, 상기 인간 모델의 특정한 포인트는 상기 인간의 머리에 대응되는 상기 인간 모델의 상부, 상기 인간의 발에 대응되는 상기 인간 모델의 하부, 상기 인간의 중심에 대응되는 상기 인간 모델의 형상의 중심(centroid)이 될 수 있다. 상기 인간 모델의 나머지는 상기 인간 모델의 특정한 포인트 및 상기 인간 모델의 나머지 사이의 고정된 관계에 기초하여 상기 연관된 식별 좌표 및 상기 인간 모델의 크기에 대하여 상기 2D 비디오 이미지 공간으로 맵핑될 수 있다. 예로서, 상기 인간 모델이 원형인 것으로 가정한다. 상기 2D 비디오 이미지 공간 내부의 각 픽셀들 위하여, 대응되는 원의 중심이 맵핑되고(예를 들면, 상기 2D 비디오 이미지 공간의 (x, y) 좌표들과 연관되어), 여기서 상기 원형의 형상의 나머지가 상기 원의 대응되는 크기(및 이의 중심에 대한 원의 알려진 관계)를 고려하여 상기 2D 비디오 이미지 공간으로 맵핑된다. 3차원 실세계에의 상기 인간의 특정한 포인트의 위치(상기 인간의 머리의 상부, 상기 인간의 발의 하부, 상기 인간의 중심과 같은)는 상기 2차원 비디오 이미지 내의 위치에 대해 특별한 관련성을 가질 수 있으며, 이에 따라ㅓ 상기 2차원 비디오 이미지 내의 상기 인간의 이러한 특정한 위치의 존재는 상기 3차원 실세계 내의 인간의 위치를 결정하는 데 이용될 수 있다.

[0051] 범용 인간 모델 모듈(303)은 또한 상기 2D 이미지 공간 내의 위치를 각기 확인하기 위해 상기 인간 모델의 크기를 결정할 수 있다. 상기 인간 모델의 크기는 상기 비디오 감시 시스템(101)의 보정으로부터 수득될 수 있다. 예를 들면, 상기 비디오 감시 시스템(101)이 보정 목적들을 위해 비디오를 취하는 동안 알려진 크기의 보정 모델은 모니터되는 지역 주위에서 이동할 수 있다. 상기 보정 모델은 상기 모니터되는 지역 주위를 걷고 있는 알려진 키의 사람이 될 수 있다. 보정 동안에, 상기 시스템은 전경 블롭으로서 상기 비디오 내의 보정 모델을 확인할 수 있고, 상기 전경 블롭이 소정의 크기(예를 들면, 소정의 높이)에 대응되는 것을 인식할(예를 들면, 상기 보정 모델의 크기에 관하여 상기 비디오 감시 시스템(101)에 제공되는 보정 정보에 접근함에 의해) 수 있다. 여기서, 상기 모정 모델이 비디오 보정 동안에 모니터되는 지역을 통해 이동함에 따라, 상기 비디오 이미지 내의 다양한 위치들에 대하여, 상기 시스템이 상기 보정 모델의 알려진 크기를 상기 2D 비디오 이미지 내의 크기와 연관시킬 수 있다. 예를 들면, 상기 보정 모델의 중심이 위치(x1, y1)에 있을 때, 상기 보정 모델의 높이는 15의 픽셀들이 될 수 있다(또는 일부 다른 측정에서 측정될 수 있다). 상기 보정 모델의 중심이 위치(x2, y2)에 있을 때, 상기 보정 모델은 높이가 27의 픽셀들이 될 수 있다. 따라서, 상기 비디오 감시 시스템(101)은 상기 2D 비디오 이미지 크기를 상기 보정 모델의 알려진 크기(예를 들면, 높이)와 연관시킴에 의해 상기 2D 비디오 이미지 내의 특정한 위치들(예를 들면, (x, y) 좌표)에서 상기 2D 비디오 이미지의 치수들을 상기 실세계의 크기들(예를 들면, 높이들)과 연관시킬 수 있다. 실세계 크기들 및 상기 2D 이미지 내의 특정한 위치들(예를 들면, (x, y) 좌표들)에서의 상기 2D 비디오 이미지의 치수들 사이의 알려진 보정(이러한 보정들 통해 얻어진)

에 기초하여, 상기 2D 비디오 이미지 공간 내의 인간 모델의 2D 크기가 실제 3D 세계 내의 평균 인간 크기에 대응되도록 상기 2D 비디오 이미지 내의 각각의 다양한 위치들((x, y) 좌표들)에 대해 계산될 수 있다

[0052] 보정 과정들의 예들에 대하여, Lipton 등에게 허여된 미국 특허 제7,932,923호 및 Zhang 등에게 허여된 미국 특허 제7,801,330호를 참조하기 바라며, 이들의 내용들은 여기에 참고로 포함된다. 일반적으로, 카메라 높이(H), 카메라 필드의 화각들(view angles)(Θ_H, Θ_V) 및 카메라 틸트(tilt) 각도(α) 그리고 객체의 외부 경계들(예를 들면, 사람의 상부 및 하부)에서 탐지되는 바와 같이 다른 정보와 같이 보정 과정을 거쳐 입력되거나 수득되는 변수들을 이용하여, 상기 카메라 시스템은 확인 목적들을 위해 일반적으로 상기 실세계 크기 및 객체의 형상을 결정할 수 있다.

[0053] 인간 기반의 카메라 보정 모델(304)은 상기 비디오 이미지 공간 내의 적절한 대응되는 위치들과 함께 상기 범용 인간 모델 모듈(303)로부터의 적절한 크기로 상기 인간 모델을 수용하고 저장할 수 있다. 이들 인간 모델들 및 대응되는 위치들은 룩업(look-up) 테이블 내에 저장될 수 있다. 예를 들면, 상기 비디오 이미지 공간 내부 및 외부의 각각의 복수의 (x, y) 좌표들은 대응되는 인간 모델을 확인하기 위해 사용될 수 있다. 예를 들면, 상기 (x, y) 식별 좌표가 상기 인간 모델의 중심에 대응될 때, 위치 (x1, y1)을 중심으로 하는 비디오 이미지 내의 인간 객체의 존재의 추정에서, 상기 인간 기반의 카메라 보정 모델(304)의 룩업 테이블은 입력으로서 위치 (x1, y1)을 수신할 수 있고, 대응되는 인간 모델(상기 2D 이미지 공간 내의 이의 크기 및 위치를 포함하여)을 제공할 수 있다. 예를 들면, 상기 출력은 상기 2D 이미지 공간 내의 경계를 포함할 수 있거나, 상기 대응되는 인간 모델을 서술하도록 상기 이미지 공간 내의 픽셀들의 완전한 세트(예를 들면, 모든 픽셀들의 (x, y) 좌표들)를 포함할 수 있다.

[0054] 도 3b는 각기 상기 2차원 비디오 이미지에 대한 다른 위치에 대응되는 몇몇 인간 모델들이 2차원 비디오 이미지를 차지하는 예를 나타낸다. 예시한 바와 같이, 네 인간 모델들(320a, 320b, 320c, 320d)이 상기 2차원 비디오 이미지에 대한 다른 (x, y) 식별 좌표들과 연관된다. 인간 모델(320a)은 가장 작으며, 3차원 실세계 내의 상기 비디오 소스로부터 가장 멀리 떨어진 위치에 대응된다. 인간 모델들(320b, 320c, 320d)은 상기 비디오 소스에 계속적으로 보다 가까운 상기 3차원 실세계 내의 위치들에 대응된다. 상기 인간 모델들(320a, 320b, 320c, 320d)은 모두 동일한 전체 인간 형상 모델로부터 유도될 수 있다. 그러나, 상기 전체 인간 형상 모델의 일부만이 특정한 위치들에서 상기 2차원 비디오 이미지를 차지하는 점도 추정될 수 있다. 여기서, 상기 전체 인간 형상 모델이 상기 2차원 비디오 이미지 공간(330)을 부분적으로만 차지하는 인간 형상들(320c, 320d)에 대응되고; 인간 모델(320c)이 상기 전체 인간 형상 모델의 발가락 및 머리 결합으로 추정되고, 여기서 인간 모델(320d)이 상기 전체 인간 형상 모델의 머리 부분에만 대응되는 점이 추정된다.

[0055] 각 인간 모델(320a, 320b, 320c, 320d)은 상기 2차원 비디오 이미지에 대한 (x, y) 식별 좌표와 연관된다. 이러한 예에서, 인간 모델들(320a, 320b 320c)의 식별 좌표들은 상기 인간 모델의 중심에 대응된다. 추정된 형상들(320a, 320b, 320c)과 연관되는 (x, y) 식별 좌표는 각기 321a, 321b 및 321c이며, 상기 비디오 이미지의 (x, y) 좌표 내에 위치한다. 추정된 형상(320d)과 연관되는 (x, y) 식별 좌표는 상기 비디오 이미지의 (x, y) 좌표들의 외부에 위치한다. 즉, 이러한 예에서, 320d와 연관된 상기 인간 형상 모델의 중심이 상기 비디오 이미지 아래에 위치하며, 이에 따라 이의 식별 (x, y) 좌표가 음의 y-축 값을 가지고, 이는 이러한 예에서 상기 비디오 이미지(도 3b에는 도시되지 않음)의 좌표들의 외부에 있다. 보정들을 용이하기 하기 위하여, 상기 (x, y) 식별 좌표들이 픽셀 단위로 증가될 수 있으므로 식별 좌표들(321a, 321b, 321c)도 상기 비디오 이미지의 픽셀들을 확인한다.

[0056] 도 3b는 설명의 편의의 목적들을 위하여 네 해당 식별 좌표들과 연관된 네 인간 모델들만을 예시한다. 그러나, 인간 기반의 카메라 보정 모델(304)은 보다 큰 숫자의 (x, y) 식별 좌표들을 위한 인간 모델들을 저장할 수 있으며, 이들의 몇몇은 인간 모델들이 서로 중첩될 수 있게 한다. 도 3c는 각기 대응되는 인간 모델(320)과 연관되는 (x, y) 식별 좌표들(321)의 단일 열을 예시한다. 예시의 편의를 위하여, 단일 열만이 예시되지만, 인간 모델들이 (x, y) 식별 좌표들의 복수의 열들에 대해 제공될 수 있으며, 이들은 상기 이미지 공간(330) 상부의 x 및 y 방향들로 규칙적으로 분산될 수 있다. 논의된 바와 같이, 상기 형상들의 크기는 다른 위치들(비록 이들이 도 3c에서 동일한 크기를 가지는 것으로 도시되지만)에 대해 상이할 수 있다. 예를 들면, 인간 기반의 카메라 보정 모델(304)은 상기 2D 이미지 공간(330)의 (x, y) 식별 좌표로서 상기 2D 이미지 공간(330) 내의 모든 픽셀에 대해서 뿐만 아니라 적어도 부분적으로 상기 2D 이미지 공간(330) 내에 위치하는 인간 모델과 연관된 상기 2D 이미지 공간(330) 외부의 (x, y) 좌표에 대해서 인간 형상을 저장할 수 있다. 예를 들면, 상기 비디오 이미지 공간(330) 내의 모든 (x, y) 픽셀 좌표들을 위하여, 인간 기반의 카메라 보정 모델(304)은 상기 인간 모델의

중심이 상기 비디오 이미지의 비디오 이미지 공간(330) 내의 이러한 (x, y) 식별 좌표에 위치할 때에 인간에 의해 점유될 것으로 예상되는 상기 비디오 이미지 공간(330) 내의 부분 공간(sub-space)의 (x, y) 식별 좌표 및 연관된 인간 모델(경계나 픽셀들의 세트를 포함할 수 있다)을 저장할 수 있다. 상기 (x, y) 식별 좌표들은 또한 상기 비디오 이미지 공간(330) 내의 부분 공간 내부의 인간 모델과 연관된 상기 비디오 이미지 공간(330) 외부의 모든 (x, y) 식별 좌표들을 포함할 수 있다(즉, 상기 전체 인간 모델의 일부가 상기 비디오 이미지 공간(330)의 부분 공간 내에 위치할 수 있다). 일부 상황들에 대하여, 앞서 언급한 부분 공간은 상기 전체 비디오 이미지 공간(330)(인간이 상기 비디오 이미지를 완전치 차지하도록 위치할 때의 추정에 대응되는)을 포함할 수 있다. 상기 인간 기반의 카메라 보정 모델(304)은 상기 (x, y) 식별 좌표들 및 연관된 인간 모델을 룩업 테이블로서 저장할 수 있다. 상기 전체 인간 형상 모델이 이러한 예에서 상기 인간 모델의 (x, y) 식별 좌표들에 대응되지만, 상기 인간 형상 모델의 다른 식별 포인트들(예를 들면, 눈, 코, 머리의 중심, 머리의 상부, 발가락, 발의 저부 등)이 이용될 수 있다.

[0057] 인간 확률 맵 연산(human probability map computation) 모듈(305)은 각 이미지 픽셀 위치에 대해서와 같이 상기 2차원 비디오 이미지 내의 각각의 복수의 위치들에 대한 인간 타겟 확률을 계산하도록 상기 전경 블롭 탐지 모듈(105) 및 상기 인간 기반의 카메라 보정 모델(304)로부터의 이들의 대응되는 식별 좌표 출력과 함께 상기 인간 모델들에 의한 비디오 이미지 출력의 특정한 프레임의 전경 블롭 세트를 이용한다. 상기 복수의 계산될 확률들은 확률 맵을 생성하도록 상기 복수의 위치들과 연관될 수 있다. 상기 복수의 위치들은 상기 인간 모델들의 (x, y) 식별 좌표들과 동일할 수 있다.

[0058] 각 (x, y) 식별 좌표를 위하여, 비디오 이미지 내의 인간 객체의 존재의 대응되는 확률을 결정하기 위해 계산이 수행된다. 상기 (x, y) 식별 좌표들이 상기 비디오 이미지의 픽셀들과 일대일의 관련성을 가질 때, 그러면 확률 계산은 상기 비디오 이미지의 각각의 픽셀들에 대해 수행된다. 예를 들면, 각 이미지 픽셀을 위하여, 대응되는 인간 가능성은 그 이미지 중심이 고려되는 픽셀 상에 있는 인간 타겟의 존재의 가능성으로 계산될 수 있다. 확률 맵은 각 (x, y) 식별 좌표에 대한 각각의 확률 계산들에의 맵핑을 생성할 수 있다. 상기 확률 맵은 각 (x, y) 좌표(입력으로서)를 상기 연관되는 계산된 확률과 연관시키는 룩업 테이블 내에 저장될 수 있다. 이러한 룩업 테이블은 상기 인간 기반의 카메라 보정 모델 모듈(304)의 룩업 테이블(엔트리로서 인간 모델들을 저장하는)과 동일할 수 있거나, 제2의 별개의 룩업 테이블이 될 수 있다.

[0059] 상술한 바와 같이, 식별 좌표들은 상기 비디오 이미지 공간 외부에 있을 수 있으며, 이에 따라 상기 비디오 이미지(이들 식별 좌표들과 연관된 상기 이미지 공간(상기 인간 모델) 내에 있는 대응되는 전체 인간 2D 모델의 일부에 관하여) 내의 상기 인간 객체의 존재의 대응되는 확률을 결정하도록 계산들이 수행될 수 있다. 예를 들면, 2D 전체 인간 모델의 중심이 상기 식별 좌표들에 대응되는 경우, 이는 상기 비디오 이미지 공간 외부에 위치할 수 있지만, 상기 전체 인간 모델의 일부인 상기 비디오 이미지 공간 내의 2D 인간 모델에 대응될 수 있다. 예를 들면, 비록 이러한 전체 인간 모델의 중심(예를 들면, 상기 전체 인간 모델의 배꼽 근처)이 상기 이미지 공간(대응되는 어깨들/머리 2D 인간 모델을 확인하는 데 이용되는 상기 식별 좌표들에 대응되는 중심) 외부에 있을 수 있지만, 전체 인간 모델의 어께들과 머리가 상기 2D 인간 모델(상기 어께들과 머리가 상기 이미지 공간 내에 있는)을 구성할 수 있다. 일부 예들에서, 상기 전체 인간 2D 모델의 특정 퍼센티지는 수행되는(또는 고려되는) 확률 계산을 위해 상기 이미지 공간 내에 있어야 한다. 예를 들면, 상기 전체 인간 2D 모델의 10% 이하 또는 20% 이하가 상기 이미지 공간 내에 있을 때(또는 상기 인간 모델이 상기 전체 인간 2D 모델의 10% 이하이거나 20% 이하일 때), 상기 식별 좌표들과 연관된 확률 값이 영(zero)으로 설정되어야 하거나 무시되어야 한다. 일부 예들에서, 상기 전체 인간 2D 모델의 40% 이하가 상기 이미지 공간 내에 있을 때, 상기 식별 좌표들과 연관된 확률 값이 영으로 설정되어야 한다.

[0060] 각 (x, y) 식별 좌표에 대한 확률 계산은 상기 대응되는 (x, y) 식별 좌표 및 상기 전경 블롭 세트와 연관된 상기 인간 모델의 리콜(recall)일 수 있다. 예를 들면, 각 (x, y) 식별 좌표를 위한 확률 계산은 상기 대응되는 (x, y) 식별 좌표와 연관된 상기 인간 모델 내의 상기 인체 픽셀들 및 상기 인체 경계 픽셀들의 리콜일 수 있다. 상기 대응되는 (x, y) 식별 좌표와 연관된 상기 인간 모델은 상기 인간 기반의 카메라 보정 모델 모듈(304)(예를 들면, 상기 모듈(304)의 룩업 테이블 내에 저장된)로부터 출력될 수 있다. 상기 전경 블롭 세트는 상기 전경 탐지 블롭 모듈(105)로부터 출력될 수 있다. 상기 전경 블롭 세트와 함께 추정된 형상의 리콜은 상기 전경 블롭 세트와 중첩되는 인간 모델 면적의 비율로서 계산될 수 있다. 특정 쓰레시홀드를 초과하지 않는 확률 계산들은 무시될 수 있다. 예를 들면, 0.4(0 내지 1의 범위에서) 이하의 계산된 확률들은 이러한 위치를 중심으로 하는 인간 타겟이 존재하지 않는 것을 나타낼 수 있다. 리콜 계산 이외의 계산들은 각각의 상기 복수의 추정된 형상들에 대응되는 상기 비디오 이미지 내의 인간 객체의 존재의 가능성을 결정하도록 수행될 수 있다. 상기

계산된 확률들이 추정치들인 점이 이해될 것이다. 따라서, 1(0 내지 1의 범위에서)의 계산된 확률이 관련된 대응되는 위치에서 인간의 존재의 절대적인 확실성을 나타내지는 않는다.

[0061] 도 3d는 도 3a의 시스템에 의해 구현될 수 있는 인간 확률 맵을 계산하기 위한 예시적인 방법을 나타낸다. 단계 S340에서, 304내의 상기 보정된 카메라 모델이 상기 2D 이미지 공간의 이미지 평면을 상기 실세계 그라운드 평면상으로 맵핑하도록 사용될 수 있다. 단계 S342에서, 인간 모델이 상기 2D 이미지 공간 내의 N의 위치들에 대해 수득될 수 있다(N은 2와 같거나 큰 정수이다). 상기 보정된 카메라 모델(304)은 상기 2D 이미지 공간 내의 모든 이미지 픽셀 위치를 위한 인간 모델로서 대응되는 볼록 헐(convex hull) 형상의 인간 모델을 수득하는 데 사용될 수 있다. 각각의 상기 인간 모델들은 상기 2D 이미지 공간 내의 식별 좌표와 연관될 수 있다. 예를 들면, 상기 인간 모델의 인간 중심 포인트는 상기 식별 좌표에 대한 맵핑을 수행할 때에 기준 포인트로서 이용될 수 있다. 상기 2D 이미지 공간의 식별 좌표가 상기 이미지 공간 내의 인간의 중심이라고 가정하면, 상기 실세계 그라운드 평면상의 이의 대응되는 물리적인 발자국 위치는 상기 보정된 카메라 모델(예를 들면, 도 5에 도시한 바와 같은)을 통해 계산될 수 있다. 범용 3D(예를 들면, 다중 실린더) 인간 모델이 이후에 이러한 발자국 위치상에 놓인다. 상기 3D 모델의 크기는 이미 수득된 보정 데이터에 대응될 수 있다. 상기 범용 3D 인간 모델은 상기 2D 이미지 공간 내에 인간 모델을 얻기 위해 상기 2D 이미지 평면상으로 투영되거나 맵핑될 수 있다. 예를 들면, 3D 다중 실린더 인간 모델의 투영은 상기 연관된 식별 좌표(예를 들면, 고려되는 이미지 포인트)에서 중심을 갖는 상기 이미지 인간 모델로서 대응되는 2D 이미지의 볼록 헐을 형성하도록 이용될 수 있다. 이러한 방식으로 모든 유효한 이미지 픽셀은 상기 이미지 위치에서 대략적인 인간 크기 및 형상을 나타내는 대응되는 볼록한 영역 형상의 인간 모델(상기 인간 모델로서)을 가질 수 있다. 계산 비용을 감소시키기 위하여, 상기 볼록한 영역 형상의 인간 모델들은 상기 시스템의 초기화에서 미리 계산될 수 있고, 상기 인간 볼록 모델의 사각형의 바운딩 박스(bounding box)는 적분 영상(integral image)을 이용하여 대략적인 인간 리콜 비율을 수득하는 데 이용될 수 있다. 단계 S344에서, 상기 전경 블롭 세트가 비디오 이미지로부터 추출될 수 있다. 상기 전경 블롭 세트는 모듈(301)에 의해 추출된 상기 인간 전경 픽셀들 및/또는 모듈(302)에 의해 추출된 상기 인간 경계 픽셀들을 이용하여 탐지된 하나 또는 그 이상의 전경 블롭들을 포함할 수 있다. 단계 S346에서, 각각의 N의 위치들에 대하여, 이러한 위치에서 인간의 존재의 확률이 확률 맵을 수득하기 위해 계산된다. 상기 인간 확률 측정은 상기 이미지 인간 볼록 모델 내의 충분한 인간 경계 픽셀들이 있는 것으로 정해진 인간 리콜 비율로서 정의될 수 있다. 이러한 예에서의 인간 리콜 비율은 이러한 인간 볼록 모델의 전체 면적에 대한 이미지 인간 볼록 모델 내의 301에서 계산된 인간 전경 픽셀들의 숫자이다. 도 3d의 프로세스의 단계들의 순서는 도시된 바와는 다른 순서로 수행될 수도 있다. 예를 들면, 단계 344는 단계들 340 및 342의 하나 또는 모두 이전에 수행될 수 있다.

[0062] 도 3a를 참조하면, 305에서 계산된 인간 확률 맵에 기초하여, 인간 타겟 추정(human target estimation) 모듈(306)은 상기 비디오 이미지 및 이들의 위치들 내의 인간 모델들(예를 들면, 인간 객체들)의 최적의 숫자를 찾을 수 있다. 전역 최적화(global optimization method) 방법이 인간 모델들 및 이들의 위치들의 최적의 숫자를 찾는 데 이용될 수 있다. $m(m_1,\cdots, m_M)$이 상기 이미지 공간 내의 모든 잠재적인 인간 모델들로부터의 M 세트의 인간 모델들을 나타내는 경우, 목적은 최적의 세트 n*을 찾는 것이므로 평가 함수(criterion function) f(n*)은 전체 최대에 도달한다. 즉, 목적은 다음을 찾는 것이다.

[0063] $$\underset{n\in m}{\operatorname{argmax}} f(n)$$

[0064] 여기서, n은 상기 이미지 공간 내의 복수의 인간 모델들의 특정한 세트이고, f(n)은 인간 모델들의 이러한 세트에 대해 계산된 함수이다.

[0065] 다음에 더 설명하는 바와 같이, 상기 함수 f(n)은 인간 모델들의 몇몇의 선택된 세트들 각각에 대해 계산되며, 각각의 세트는 상기 확률 맵으로부터 m_i 위치들을 선택한다(m_i 위치들은 각 패스(pass)에 대해 선택되고, 여기서 상기 숫자 m_i는 각각의 이들 패스들에 대해 다르다). 인간 모델들의 각 세트는 각 패스에 대해 변경되는 위치들을 선택하는 데 이용되는 특정한 제한적인 기준들로 상기 확률 맵의 패스(또는 스캔)와 함께 선택될 수 있다. 여기서, 상기 함수 f(n)은 다음과 같이 정의된다.

[0066] $$f(n) = w_R * R(n) + w_P * P(n) - w_O * O(n)$$

[0067] 여기서, R은 상기 인간 리콜 비율이며, 이는 n의 선택된 인간 모델들의 그룹의 전체 면적에 대한 상기 인간 전

경 면적의 퍼센티지로서 정의되고; P는 인간 정확성이며, 이는 상기 n의 선택된 인간 모델들의 그룹과 중첩되는 상기 전경 면적의 퍼센티지이고; O는 인간 중첩 비율이며, 이는 모든 n의 선택된 인간 모델들에 의해 차지되는 면적에 대해 서로의 상기 n의 선택된 인간 모델들의 임의의 것의 중첩의 면적의 비율이고; w_R, w_P 및 w_O는 가중치들(weights)이다. 너무 많은 인간 중첩 없이 상기 전경 영역(전경 블롭 세트) 및 상기 인간 모델들(m의 인간 모델들의 세트)의 결합 사이의 가장 우수한 매칭(matching)을 찾는 것이 유리할 수 있다. 실제로, 어떻게 전술한 세 가중치들을 결정할 것인가는 탐지 결과들에 상당히 중요한 영향을 미친다. 예를 들면, 보다 큰 가중치가 상기 인간 중첩 비율에 가해질 경우, 이는 보다 적은 인간 계수를 가져올 수 있다.

[0068] 각각의 상기 m_i의 선택된 인간 모델들은 상기 인간 확률 맵 연산 모듈(305)에 의한 상기 확률 맵 출력에 대한 참조에 의하여 선택될 수 있다. 몇몇의 패스들이 계산 f(n)을 수행하도록 이루어질 수 있으며, 각 패스는 상기 범용 인간 모델 모듈(303)에 의해 제공되고 인간 기반의 카메라 보정 모델(304)(예를 들면, 룩업 테이블 내의) 내의 (x, y) 식별 좌표와 연관되는 상기 2D 인간 모델들로부터 m_i의 인간 모델들의 서브세트를 선택한다. 언급한 바와 같이, m_i의 값은 각각의 이들 패스들에 대해 다를 수 있다. 상기 인간 모델들의 선택 기준들이 각 패스에 대해 다를 수 있으므로 다른 인간 모델들이 다른 패스들에 대해 선택된다(또한, 가능한대로, 다른 숫자의 m_i의 인간 모델들이 다른 패스들에 대해 선택된다). 선택 기준들은 상기 확률 맵에 의해 설정되는 바와 같은 확률 쓰레시홀드 P_{th}와 연관된 상기 선택된 인간 모델을 요구하는 것을 포함할 수 있다. 선택 기준들은 또한 임의의 미리 선택된 2D 인간 모델들로부터 최소 거리 D_{min}으로 떨어진 다음의 선택된 2D 인간 모델을 포함할 수 있다. 상기 최소 거리 D_{min}은 상기 실세계의 그라운드 평면상의 거리일 수 있다. 예를 들면, 상기 2D 인간 모델들의 중심들이 상기 3D 실세계 내의 위치들로 맵핑되거나 해석될 수 있으며, 이들 사이의 거리들이 계산될 수 있다. 상기 최소 거리들 D_{min}은 상기 2D 이미지 평면 내에서 계산될 수 있지만, 상기 2D 이미지 평면 내의 거리들이 대응되는 3D 위치들을 반영할 수 있으므로, 상기 비디오 이미지 소스 부근의 인간 모델들에 대하여, 보다 큰 분리가 보다 먼 인간 모델들에 대한 경우보다도 상기 2D 이미지 평면 내에 요구될 수 있다.

[0069] 일부 예시적인 실시예들에 있어서, 상기 확률 맵의 하나 또는 그 이상의 빠른 원-패스 스캐닝(one-pass scanning)이 인간 계수 및 대응되는 위치들을 결정하는 데 이용된다. 도 3e는 상기 비디오 이미지 내의 인간 모델들의 최적의 숫자를 발견하는 과정의 부분으로서 상기 확률 맵의 단일 패스를 수행하는 방법을 예시한다. 도 3e의 방법은 인간 타겟 추정 모듈(306)에 의해 구현될 수 있다. 단계 S350에서, 확률 맵은 국지적 최대(특정 선택 기준들에 의해 부여될 수 있다)를 찾기 위해 스캔된다. 상기 확률 맵은 상기 비디오 소스에 가장 가까운 실세계 내의 위치에 대응되는 이용 가능한 선택되지 않은 국지적 최대를 위치시키도록 스캔될 수 있다. 상기 확률 맵의 하부는 상기 비디오 이미지의 하부에 대응될 수 있다. 많은 구현예에서, 감시 기능을 수행하는 비디오 카메라는 상기 모니터되는 지역 내의 인간들의 머리 레벨들 보다 높은 위치에 장착될 수 있다. 따라서, 상기 비디오 이미지의 하부는 상기 비디오 소스에 가장 가까운 위치에 대응될 수 있다. 이러한 예에서 하부에서 상부까지 상기 확률 맵을 스캐닝하는 것은 인간 모델들의 선택이 상기 비디오 이미지 내의 가려진 객체에 대응될 가능성이 적어지게 한다.

[0070] 상기 확률 맵은 상기 이미지 공간 내의 각각의 상기 복수의 위치들에 대해 상기 미리 계산된 확률들의 국지적 최대(상기 확률 맵에 저장된)를 나타내는 국지적 최대 포인트를 찾기 위해 하부부터 상부까지 스캔될 수 있다. 상기 국지적 최대는 바로 옆에 이웃하는 (x, y) 식별 좌표들(예를 들면, 바로 옆에 이웃하는 픽셀들)의 각각의 확률 값들 보다 높은 확률 값을 갖는 (x, y) 식별 좌표(예를 들면, 픽셀)일 수 있다. 국지적 최대 포인트가 발견되면, 이의 식별 좌표로서 이러한 국지적 최대 포인트와 연관된 인간 모델이 m_i의 인간 모델들의 세트의 하나로 단계 S352에서 선택된다. 단계 S354에서, 이러한 선택된 모델의 내부 영역 내의(예를 들면, 상기 2D 인간 모델의 경계 내에 있는) 모든 픽셀들 및 이러한 선택된 모델로부터 떨어진 최소 거리 D_{min}에 대응되는 픽셀들(예를 들면, 상기 실세계의 그라운드 평면상의 최소 거리를 나타내는 상기 비디오 이미지 내의 픽셀들)은 이러한 패스 내에서 다른 고려로부터 배제된다(또한 이러한 패스를 위한 상기 확률 맵으로부터 일시적으로 제거될 수 있다). 이러한 예에서, 픽셀들이 상기 인간 모델들의 식별 좌표들에 대응되며, 이러한 설명이 픽셀 위치들이 아닌 식별 좌표들에 동등하게 적용 가능한 점에 유의한다. 일부 예들에서, 상기 비디오 이미지 자체는 이러한 단계에서 더 분석될 필요는 없으며, 상기 픽셀들은 상기 확률 맵으로부터의 이들의 일시적 제거에 의해 다른 고려로부터 간단히 제외될 수 있다. 상기 확률 맵은 상기 확률 쓰레시홀드 P_{th} 보다 크고 제외되지 않았던 픽셀들에 연관되는 상기 인간 확률 맵의 확률들의 다른 국지적 최대 포인트를 선택하도록 다시 스캔된다. 단계 S356에서, 임의의

유효한 픽셀들이 고려되었는지가 결정된다. 즉, 상기 확률 맵은 상기 선택 기준들에 의해 제외되지 않았거나 상기 확률 맵의 이러한 스캔 내의 다른 인간 모델들의 선택에 의해 제외되지 않았던 경우의 값들에 대해 리뷰된다. 상기 확률 맵의 스캔은 모든 유효한 픽셀들이 고려되고 상기 맵으로부터 제거될 때까지 계속된다. 따라서, m_i의 인간 모델들이 상기 확률 맵의 이러한 스캔으로 선택될 수 있다. 이와 같은 패스를 위하여, 상기 함수 $f(m_i)$가 이러한 m_i의 인간 모델들의 세트에 대하여 계산된다.

[0071] 상기 확률 맵의 추가적인 스캔들은 선택 기준들의 다른 세트를 갖는 각 온-패스(on-pass) 스캔으로 수행될 수 있다. 도 3f는 비디오 이미지 내의 인간 모델들의 최적의 숫자를 찾는 것에 대해서 상기 확률 맵의 복수의 패스들을 수행하는 방법을 예시한다. 도 3f의 방법은 인간 타겟 추정 모듈(306)에 의해 구현될 수 있다. 여기서, D_{min}(최소 거리) 및 P_{th}(확률 쓰레시홀드)의 적어도 하나의 값이 각 스캔에 대해 다를 수 있다. 단계 S360에서, 상기 선택 기준들은 특정한 온-패스 스캔을 위해 설정된다. 상기 선택 기준들의 얼마나 많은 변경들(및 이에 따라 얼마나 많은 스캔들)이 원하는 정확도와 계산의 부담을 고려하여 한 건 한 건을 기초로 하여 결정될 수 있다. 단계 S362에서, 상기 확률 맵의 스캔은 상기 선택 기준들에 따라 m의 인간 모델들의 세트를 선택하도록 수행된다. 상기 값 m은 0과 같거나 그 이상인 정수이며 각 선택에 대해(예를 들면, 단계 S362를 수행하는 도 3f의 각 루프에 대해) 다를 수 있다. 단계 S362는 도 3e의 방법에 대응될 수 있다. 단계 S364에서, 평가 함수가 상기 선택된 m_i의 인간 모델들에 대해 계산된다. 예를 들면, 대응되는 $f(m_i)$가 이러한 스캔에서 선택된 상기 m_i의 인간 모델들에 대해 계산된다. 추가적인 스캔들이 새로운 선택 기준들(S366)로 수행될 수 있다. 상기 확률 맵의 모든 스캔들이 완료될 때, n∈스캔들의 집합의 $\{m_1,\cdots, m_M\}$인 f(n)의 최대가 결정된다. 이러한 최대값에 대응되는 인간 모델들의 세트는 상기 비디오 이미지 내의 인간 객체들에 대응되도록 결정된다(S368). 상기 비디오 이미지 내의 인간 객체들을 나타내도록 결정된 인간 모델들의 상기 (x, y) 식별 좌표들(예를 들면, 픽셀 위치들)을 이용하여, 상기 그라운드 평면상의 실세계 위치가 결정될 수 있다.

[0072] 선택적인 실시예에 있어서, m이 상기 이미지 공간 내의 모든 잠재적인 인간 모델들로부터의 인간 모델들의 세트를 나타낼 경우, 목적은 최적 세트 m*을 찾는 것이 될 수 있으므로 평가 함수 g(m*)은 전체 최대에 도달한다. 즉, 목적은 다음의 최대를 찾는 것이다.

$$g(m) = \sum_{n=1}^{m} f(n)$$

[0073]

[0074] 여기서, n은 상기 이미지 공간 내의 복수의 인간 모델들의 하나이고, m은 많은 선택된 인간 모델들이며(다른 합산 계산에 대해 변화될 수 있다), f(n)은 모델들의 그룹 보다는 각각의 상기 m의 인간 모델들에 대해 계산된 함수이다.

[0075] 이 경우, 상기 함수 f(n)은 다음과 같이 정의된다.

$$f(n) = w_R * R(n) + w_P * P(n) - w_O * O(n)$$

[0076]

[0077] 여기서, R은 상기 인간 리콜 비율이고, 이는 상기 선택된 인간 모델들의 전체 면적에 대한 상기 인간 전경 면적의 퍼센티지로 정의되며; P는 인간 정확성이고, 이는 상기 선택된 인간 모델들과 중첩되는 상기 전경 면적의 퍼센티지이며; O는 선택된 n번째 인간 모델과 1번째 내지 n-1번째 인간 모델들에 의해 점유되는 면적들[Σf(n)의 계산에서 현재 패스 내에 미리 선택된 인간 모델들에 의해 점유되는 면적들]의 중첩이며; w_R, w_P 및 w_O는 가중치들이다. 전술한 상기 확률 맵을 스캐닝하는 각각의 패스들은 각 패스에 대한 상기 확률 맵의 국지적 최대들의 선택에서 상기 선택 기준들 상의 다른 제한들로 Σf(n)의 계산과 연관될 수 있다.

[0078] 도 6a, 도 6b 및 도 6c는 일 예에 따른 비디오 감시 시스템(101)의 탐지 결과들을 나타낸다. 하나의 출력 프레임에 대하여, 도 6a는 상기 인체 탐지 모듈(301) 및 인간 경계 픽셀 탐지 모듈(302)의 출력이며, 여기서 상기 픽셀들(301a)은 상기 탐지된 인체 픽셀들을 나타내고, 상기 픽셀들(302a)은 상기 인간 경계 픽셀들을 나타낸다. 상기 전경 블롭 세트는 상기 탐지된 인체 픽셀들(301a) 및 상기 인간 경계 픽셀들(302a)의 결합으로서 도 6a에 나타난다. 상기 탐지된 인체 픽셀들 및 인간 경계 픽셀들은 상기 비디오 이미지 공간(330)을 한정하는 최초의 비디오 이미지 프레임에 대해 중첩된다. 이와 같은 예에서, 이러한 비디오 이미지 프레임(상기 전경 블롭 세트보다는) 내의 상기 비디오 이미지의 나머지는 상기 배경 이미지의 일부이다.

[0079] 도 6b는 도 6a로부터 계산된 인간 확률 맵을 예시한다. 이러한 예에서, 상기 인간 확률 맵은 영(0)의 확률에 대응되는 흑색 및 일(1)의 확률에 대응되는 백색을 갖는 그레이 스케일(grey scale) 상의 계산된 확률들을 나타낸다. 각각의 계산된 확률들은 대응되는 인간 모델의 식별 좌표에 대응되는 픽셀들에 대응되는 상기 이미지 공간(330) 내의 위치를 나타낸다.

[0080] 도 6c는 탐지된 인간에 대응되는 복수의 인간 모델들(320)(핑크색 볼록 형상 윤곽)을 예시하는 최종 인간 탐지 결과를 나타낸다. 각각의 이들 인간 모델들은 상기 3D 실세계 내의 탐지된 인간의 위치를 확인할 수 있고 상기 실세계의 그라운드 평면으로 맵핑될 수 있는 식별 좌표(중심과 같은)에 의해 연관될 수 있다.

[0081] 도 7a, 도 7b 및 도 7c는 상기 인간 탐지 결과들에 기초하여 인간 군중 밀도를 측정하는 실시예를 나타낸다. 도 7a는 각기 탐지된 인간에 대응되고 최초의 비디오 이미지와 중첩되는 복수의 2D 인간 모델들(320)(핑크색 볼록 헐들)을 보여주는 상기 비디오 감시 시스템(101) 탐지 결과들의 예시적인 결과를 나타낸다. 도 7b는 상기 실세계의 물리적인 그라운드 평면으로 맵핑되는 바와 같은 인간 모델(320)을 나타내고 이에 따라 상기 실세계 내의 탐지된 인간들의 위치를 확인하는 각각의 원들을 갖는 도 7a의 비디오 이미지의 상부에서 하부의 표현을 보여주는 상기 실세계의 물리적 그라운드 평면에 대한 상기 탐지된 인간들의 맵핑을 나타낸다. 상기 탐지된 인간 타겟들은 보정이 상기 보정 모델의 알려진 크기, 상기 2D 이미지 내의 위치 및 상기 이미지 공간 내의 대응되는 크기 사이의 상호관련성을 제공하였기 때문에 물리적인 그라운드 평면상으로 맵핑될 수 있다. 알려진 위치들로써, 계산들이 특정한 확인된 면적(예를 들면, 사용자에 의해 선택된) 또는 전체 장면 내의 사람들의 숫자를 세도록 수행될 수 있다. 계산들은 또한 면적당 사람들의 숫자를 결정하도록 수행될 수 있다. 각 그라운드 위치상의 실제 군중 밀도 측정들도 직접적으로 계산될 수 있다. 상기 군중 밀도 측정의 실제 정의는 실제 응용, 특히 모니터되는 군중의 크기에 의존할 수 있다. 예를 들면, 도 6a, 도 6b 및 도 6c에 도시된 시나리오들에 대하여, 상기 군중 밀도 측정으로서 2미터의 반경 내의 사람들의 숫자를 사용할 수 있다. 반면에, 도 7a, 도 7b 및 도 7c의 시나리오들에 대하여, 상기 위치의 군중 밀도는 6미터의 반경 내의 사람들의 숫자로서 정의될 수 있다. 도 7c는 보다 높은 군중 밀도를 의미하는 보다 높은 강도의 핑크색을 갖는 6미터의 반경을 사용하여 상기 군중 밀도 맵을 나타낸다.

[0082] 각 비디오 프레임에 대한 상기 군중 밀도 측정들에 기초하여, 도 1의 사건 탐지 모듈(108)의 각각의 모듈들(801, 802, 803)에 의해 탐지될 수 있는 군중 탐지, 군중 집회 및 군중 해산을 포함하여 도 8에 도시된 바와 같은 사건들과 관련된 많은 군중을 탐지할 수 있다. 도 9는 어떻게 혼잡한 지역을 정의하고 탐지하는 가의 예시적인 방법을 나타낸다. 블록 901은 군중 영역 사건을 어떻게 정의하는 가를 나타낸다. 상기 사용자는 상기 이미지 상의(예를 들면, 상기 공간 면적 내의) 관심의 대상인 영역을 먼저 선택할 수 있다. 다음에, 일부 군중 밀도 쓰레시홀드는 얼마나 많은 군중이 관심의 대상인 가를 결정하는 데 사용될 수 있다. 상기 쓰레시홀드들은 면적의 특정 반경 내의 사람들의 숫자가 될 수 있다. 히스테리시스 쓰레시홀드들(hysteresis thresholds)은 보다 양호한 수행을 위해 사용될 수 있다. 예를 들면, 3미터 반경의 면적 내의 사람들의 숫자로 상기 군중 밀도를 정의할 경우, 두 군중 밀도 쓰레시홀드들인 $T_{high}=10$ 및 $T_{low}=8$을 설정할 수 있다. 영역은 대응되는 군중 밀도가 T_{high} 보다 크거나 같을 경우에만 군중 영역으로 간주될 수 있다. 군중 영역은 상기 대응되는 군중 밀도가 T_{low} 보다 작거나 같을 경우에만 비-군중이 된다. 상기 군중 영역은 확인된 군중에 의해 정의될 수 있으며, 프레임으로부터 프레임까지 위치 및/또는 형상을 변화시킬 수 있다. 상기 군중 영역의 중심은 상기 군중 위치를 기술하는 데 이용될 수 있다. 최소 지속 쓰레시홀드(duration threshold)는 군중 영역이 상기 사건 탐지를 트리거링(triggering)하기 전에 군중으로 유지되어야 하는 최소 시간 지속을 정의할 수 있다. 새로운 비디오 프레임 입력을 위하여, 블록 902는 모든 탐지된 인간 타겟들이 군중 영역에 속하는 가를 알기 위해 조사하며, 이후에 블록 903은 이들의 상태를 업데이트하기 위해 모든 군중 영역들 점검한다. 탐지되면, 군중들 및 이들의 위치들은 상기 비디오 이미지의 프레임 단위로 추적될 수 있다. 예를 들면, 군중이 탐지되고 상기 최소 쓰레시홀드 T_{low}를 만족시키도록 계속되는 한, 상기 군중 영역과 연관된 인간 모델들은 이들이 상기 최소 군중 밀도를 만족시키는 면적 내에 남는 동안에 상기 비디오 이미지의 후속되는 프레임들로 상기 군중을 정의할 수 있다. 추가적인 인간 모델들은 이들이 상기 탐지된 군중 영역으로 이동함에 따라 탐지된 군중에 추가될 수 있다.

[0083] 도 10은 각 검출된 인간 타겟에 대한 예시적인 프로세스를 나타낸다. 블록 1001은 현재의 타겟이 현존하는 군중 영역의 내부 또는 근처에 있는 지를 점검한다. "예"일 경우, 블록 1001은 이러한 영역을 위해 사람 계산을 업데이트한다. "아니오"일 경우, 블록 1002는 상기 현재 타겟의 위치상의 군중 밀도를 계산하고, 이후에 블록 1004는 상기 군중 밀도 측정이 쓰레시홀드 T_{high} 보다 크거나 같은 지를 점검한다. "예"일 경우, 새로운 군중 영역이 현재의 타겟을 중심으로 생성된다. "아니오"일 경우, 다음 인간 타겟을 처리하도록 계속된다.

[0084] 도 11은 각 군중 영역에 대한 예시적인 프로세스를 나타낸다. 블록 1101은 상기 타겟 처리 결과들에 기초하여 영역 면적과 군중 계산을 업데이트하고; 블록 1102는 상기 밀도 계산이 여전히 사용자 정의 쓰레시홀드 보다 큰지를 점검하며; "아니오"일 경우, 상기 군중 영역이 모니터링 리스트로부터 제거된다. 블록 1104는 처리 중인 상기 군중 영역의 군중 지속이 사용자 정의 쓰레시홀드 보다 길거나 같은 지를 더 점검한다. "예"일 경우, 블록 1105는 대응되는 군중 사건이 보고되었는 지 혹은 그렇지 않았던 지를 더 점검하며, 그렇지 않았을 경우, 블록 1106은 상기 군중 사건을 보고하고 이러한 군중 영역을 "보고됨"으로 표시하는 것과 같은 행동을 수행할 것이다.

[0085] 도 12는 군중 "집회" 및 "해산" 사건들을 정의하고 탐지하는 데 이용되는 방법을 나타낸다. 여기서, "집회" 및 "해산"은 군중 집회 스팟(spot)의 형성 및 종료의 두 프로세스들을 언급한다. 이러한 예에서, 군중 집회 스팟은 매우 국지적인 정지 군중 밀도를 갖는 영역을 언급하며, 퍼레이드 내와 같이 이동하는 군중과는 다르다. 그러나, 본 발명이 이에 제한되는 것은 아니며, 이러한 방법은 또한 군중 집회 스팟들의 탐지에 적용될 수 있다. 블록 1201은 어떻게 군중 집회 스팟이 정의될 수 있는 지를 나타낸다. 상기 사용자는 먼저 상기 이미지상의 관심의 대상인 영역을 선택할 수 있다. 다음에, 일부 군중 밀도 쓰레시홀드는 얼마나 많은 군중이 관심의 대상인 지를 결정하는 데 이용된다. 상기 최소 지속 쓰레시홀드는 군중 영역이 유효 집회 스팟으로 간주되는 군중으로 유지되어야 하는 최소 시간 지속을 정의할 수 있다. 블록 1202는 상기 군중 집회 스팟들을 탐지한다. 블록 1203은 상기 탐지된 군중 집회 스팟들을 업데이트하고 모니터하며, 상기 군중 "집회" 및 "해산" 사건들을 탐지한다.

[0086] 도 13은 군중 집회 스팟을 정의하는 일 예를 나타낸다. 이는 1301로 나타낸 내부 영역 및 1302로 나타낸 외부 영역을 포함한다. 상기 두 영역들은 중심점 O, 짧은 반경 r 및 긴 반경 R에 의해 정의될 수 있다. 이러한 예에서, 상기 군중 집회 스팟은 다음 두 기준들을 충족시킬 수 있다.

[0087] ● 상기 내부 영역의 군중 밀도가 미리 정해진 쓰레시홀드 보다 크거나 같아야 한다.

[0088] ● 상기 외부 영역 내의 사람 계산이 상기 내부 영역 내의 사람 계산 보다 작아야(예를 들면, 2배, 4배, 10배 등으로 작은) 한다. 선택적으로는, 상기 외부 영역 내의 군중 밀도가 상기 내부 영역의 군중 밀도 보다 작아야(예를 들면, 2배, 4배, 10배 등으로 작은) 한다.

[0089] 상기 두 기준들은 상기 내부 영역이 군중 집회 스팟일 뿐만 아니라 대규모 군중 내의 영역인 점을 나타낼 수 있다.

[0090] 도 14a 및 도 14b는 군중 집회 스팟의 예를 나타낸다. 도 14a 및 도 14b는 각기 상기 실세계의 물리적인 그라운드 평면상으로 맵핑되는 비디오 프레임 및 탐지된 인간 타겟들을 보여준다. 비록 도 14a가 보다 많은 인간 타겟들을 가지지만, 도 14b만이 앞서 정의된 바와 같은 군중 집회 스팟을 제한한다.

[0091] 도 15는 상기 군중 집회 스팟들을 탐지하는 예시적인 방법을 나타낸다. 각각의 탐지된 인간 타겟에 대하여, 블록 1501은 이가 현존하는 군중 집회 스팟에 속하는 지를 점검한다. "예"일 경우, 이는 블록 1502 내의 대응되는 군중 집회 스팟의 현재 상태를 업데이트하는 데 이용된다. "아니오"일 경우, 블록 1503은 상기 현재의 타겟이 새로운 군중 집회 스팟의 중심인 지를 더 점검한다. "예"일 경우, 블록 1504는 다른 모니터링을 위한 새로운 군중 집회 스팟을 개시한다. "아니오"일 경우, 상기 모듈은 다음의 인간 탐지를 점검하도록 계속된다.

[0092] 도 16은 상기 군중 집회 스팟들을 업데이트하고, 군중 "집회" 및 "해산" 사건들을 탐지하는 예시적인 방법을 나타낸다. 블록 1601은 고려되는 상기 비디오 프레임 상의 새로운 인간 탐지 결과들을 이용하여 상기 군중 집회 스팟의 위치 및 면적을 업데이트한다. 블록 1602는 상기 군중 "집회" 사건이 현재의 군중 집회 스팟으로부터 검출되었는지를 점검한다. "아니오"일 경우, 블록 1603은 군중 집회 스팟이 특정 지속 동안에 성공적으로 업데이트되었는지를 점검함에 의해 상기 "집회" 사건을 탐지하도록 계속된다. 이러한 지속 쓰레시홀드는 규칙 정의 시간에서 사용자에 의해 설정될 수 있다. 군중 집회 스팟이 "집회" 사건을 발생시켰으면, 블록 1604는 상기 "해산" 사건을 탐지하기 위해 상기 집회 스팟을 더 모니터한다. 여기서, 군중 "해산" 사건은 군중 집회 스팟이 짧은 기간 내에 빈 스팟 또는 작은 밀도(예를 들면, 최소 군중 밀도 쓰레시홀드 T_{low} 아래)를 갖는 스팟으로 되면서 정의된다. 블록 1604는 군중 집회 스팟의 두 특별한 순간들인 이가 혼잡하지 않게 되는 시간 및 비거나 적게 되는 경우의 시간을 탐지한다. 이들 두 순간들 사이의 시간이 사용자 정의 쓰레시홀드 보다 짧을 경우, 군중 "해산" 사건이 탐지된다.

[0093] 도 17은 본 발명이 적용될 수 있는 다중-카메라 시스템의 예를 나타낸다. 이러한 예에서, 두 카메라들(1702, 1704)은 별도로 다른 관점들로부터 관심의 대상인 장면의 비디오 이미지들을 취한다. 여기에 기재되는 비디오

감시 시스템(101) 및 방법들은 각 카메라(1702, 1704)에 대해서 상기 변화 탐지 모듈(103), 상기 움직임 탐지 모듈(104), 전경 블롭 탐지 모듈(105), 범용 인간 모델 모듈(303), 인간 기반의 카메라 보정 모델(304) 및 인간 확률 맵 연산 모듈(305)에 대해 여기서 설명한 바와 동일할 수 있다-즉, 각 카메라는 스스로 모듈이나 이들 모듈들을 위한 모듈 기능성(회로부가 공유될 경우)을 가질 수 있다.

[0094] 각 비디오 카메라(1702, 1704)의 인간 기반의 카메라 보정 모델(304)에 의해 제공되는 각각의 이미지 공간에 대한 상기 2D 인간 모델들은 또한 상기 실세계의 물리적 그라운드 평면의 좌표와 관련될 수 있다. 예를 들면, 각 카메라를 위한 인간 기반의 카메라 보정 모델 모듈(304)에 대하여, 추가적인 엔트리가 대응되는 물리적인 그라운드 평면 좌표에 대해 수행될 수 있으며, 이에 따라 각각의 N의 인간 모델들을 동일하게 연관시킬 수 있다. 각각의 상기 카메라들(1702, 1704)을 위한 인간 확률 맵의 계산에서, 각 확률 맵의 확률들은 상기 2D 이미지 공간보다는 상기 물리적 그라운드 평면으로 맵핑될 수 있다.

[0095] 일 예에서, 인간들의 최적의 숫자를 탐지하는 인간 타겟 추정 모듈(306)은 전술한 방식으로 하나의 카메라의 제1의 확률 맵의 스캔들, 즉, 상기 서치 기준들의 제약들 하에서를 수행할 수 있다. 인간 모델들의 M의 세트들 $\boldsymbol{m}(m_1,\cdots, m_M)$에 대한 최대를 결정하기 위한 평가 함수의 계산에서, 목적은 다음을 찾는 것이다.

[0096] $$\underset{n\in m}{\operatorname{argmax}} f_1(n) + f_2(n)$$

[0097] 여기서, n은 복수의 3D 인간 모델들의 특정한 세트이고, 이는 확률들이 각각의 두 인간 확률 맵들로 맵핑되는 상기 물리적 그라운드 평면 내에 식별 좌표들을 가질 수 있다. 즉, 실세계 내의 포인트를 모델 세트를 위한 인간 모델과 연관되게 선택함에 따라, 이러한 포인트와 연관된 상기 2D 이미지 공간 인간 모델들이 $f_1(n)$을 계산하는 데 이용되는 하나의 인간 모델 및 $f_2(n)$를 계산하는 데 이용되는 다른 하나로 각 카메라 시스템에 대해 확인된다. $f_1(n)$ 및 $f_2(n)$는 여기서 설명한 함수들(각기 적당한 비디오 이미지로부터 추출되는 상기 인간 전경 블롭 세트 또는 인간 전경 면적에 대한)과 동일할 수 있다.

[0098] $$f(n) = w_R * R(n) + w_P * P(n) - w_O * O(n)$$

[0099] 여기서, (상기 비디오 이미지와 연관된 각각의 n의 선택된 2D 인간 모델들 및 이러한 비디오 이미지의 인간 전경 면적에 대하여) R은 상기 인간 리콜 비율이고, 이는 n의 선택된 인간 모델들의 그룹의 전체 면적에 대한 인간 전경 면적의 퍼센티지로 정의되며; P는 인간 정확도이고, 이는 n의 선택된 인간 모델들의 그룹과 중첩되는 전경 면적의 퍼센티지이며; O는 인간 중첩 비율이고, 이는 모든 n의 선택된 인간 모델들에 의해 점유되는 면적과 1번째 내지 n-1번째 인간 모델들에 의해 점유되는 면적들을 갖는 선택된 n번째 인간 모델[f(n)의 계산에서 현재의 패스 내에 미리 선택된 인간 모델들에 의해 점유되는 면적들]에 대한 상기 n의 선택된 인간 모델들의 서로의 중첩의 면적의 비율이며, w_R, w_P 및 w_O는 가중치들이다. 상기 가중치들이 함수들 $f_1(n)$ 및 $f_2(n)$ 사이에서 다를 수 있는 점에 유의한다. 다음의 국지적 최대의 선택에서 다른 고려를 위한 픽셀들의 제외는 미리 선택된 인간 모델의 그라운드 평면 좌표와 연관된 상기 3D 인간 모델을 해당 이미지 평면 내의 각각의 상기 두 확률 맵들로 다시 투영시킬 수 있다.

[0100] 다른 선택적인 실시예에 있어서, 단일 확률 맵이 다중 카메라들을 위해 이용될 수 있다. 도 17의 예에서, 확률 계산들은 여기서 설명한 바와 같은 각각의 상기 2D 비디오 이미지들에 대해 수행될 수 있고 각기 상기 해당 2D 이미지 평면에 대응되는 두 이미지 평면 확률 맵들을 생성할 수 있다. 상기 이미지 평면 확률 맵의 확률들은 특정 쓰레시홀드(각 이미지 평면 확률 맵에 대해 동일하거나 다를 수 있는)를 초과하지 않을 경우에 영으로 설정될 수 있다. 각 이미지 평면 확률 맵 내의 식별 좌표들은 각각의 상기 이미지 평면 확률 맵들에 대해 상기 실세계 내의 그라운드 평면 좌표로 해석될 수 있고, 각 비디오 이미지에 대한 그라운드 평면 확률 맵을 생성할 수 있다. 상기 두 이미지 평면 확률 맵들은 병합된 확률 맵을 생성하도록 동일한 그라운드 평면 좌표들을 공유하는 확률들을 곱하여 병합될 수 있다. 상기 병합된 그라운드 평면 확률 맵은 국지적인 최대들을 발견하도록 스캔될 수 있다. 각각의 발견된 국지적인 최대는 이후에 적절하게 $f_1(n)$ 또는 $f_2(n)$(상술한)를 계산하는 데 사용될 수 있는 이들의 해당 이미지 공간 내의 각각의 상기 비디오 이미지들에 대한 별도의 인간 모델들을 확인할 수 있다. 복수의 국지적인 최대들을 위해 상기 병합된 그라운드 평면 확률 맵의 다중 스캔들을 수행하는 것은 후속하는 인간 모델들(각각의 상기 비디오 이미지들에 대한 하나)을 찾고 다음을 계산하도록 수행될 수 있다.

[0101] $$f_1(n) + f_2(n)$$

[0102] 선택 제한들(최소 확률 쓰레시홀드 및 상기 3D 실세계 내의 최소 거리와 같은)은 변경될 수 있고, 새로운 스캔 패스가 m의 인간 3D 모델들(이러한 예에서 2m의 2D 인간 모델들에 대응되는)의 최적 세트를 찾도록 구현된다.

[0103] 다른 예에서, 인간들의 최적 숫자를 탐지하는 인간 타겟 추정 모듈(306)은 전술한 방식으로 하나의 카메라의 제1의 확률 맵의 스캔들을 수행, 즉 서치 기준들의 제한들 내에서 상기 제1의 확률 맵의 국지적인 최대를 조사할 수 있다. m의 인간 모델들의 세트들의 최대를 결정하기 위한 평가 함수의 계산에서, 목적은 다음의 최대를 찾는 것이다.

[0104] $$\sum_{n=1}^{m} f_1(n) + f_2(n)$$

[0105] 여기서, n은 확률들이 각각의 상기 두 인간 확률 맵들로 맵핑되는 물리적 그라운드 평면 내의 상기 식별 좌표이다. 즉, 상기 실세계 내의 포인트를 선택함에 따라, 이러한 포인트와 연관된 상기 2D 이미지 공간 인간 모델들은 $f_1(n)$을 계산하는 데 이용되는 하나의 인간 모델 및 $f_2(n)$을 계산하는 데 이용되는 다른 하나로 각 카메라 시스템에 대해 확인될 수 있다. $f_1(n)$ 및 $f_2(n)$는 전술한 함수(각기 적당한 비디오 이미지로부터 추출된 상기 인간 전경 블롭 세트 또는 인간 전경 면적에 대한)와 동일할 수 있다.

[0106] $$f(n) = w_R * R(n) + w_P * P(n) - w_O * O(n)$$

[0107] 여기서, R은 상기 인간 리콜 비율이고, 이는 선택된 인간 모델들의 그룹의 전체 면적에 대한 인간 전경 면적의 퍼센티지로 정의되며; P는 인간 정확도이고, 이는 상기 선택된 인간 모델들의 그룹과 중첩되는 전경 면적의 퍼센티지이며; O는 인간 중첩 비율이고, 이는 1번째 내지 n-1번째 인간 모델들에 의해 점유되는 면적들[$\Sigma f(n)$의 계산에서 현재의 패스 내에 미리 선택된 인간 모델들에 의해 점유되는 면적들]에 대한 상기 선택된 인간 모델들의 중첩이며, w_R, w_P 및 w_O는 가중치들이다. 상기 가중치들이 함수들 $f_1(n)$ 및 $f_2(n)$ 사이에서 다를 수 있는 점에 유의한다. 다음의 국지적 최대의 선택에서 다른 고려를 위한 픽셀들의 제외는 미리 선택된 인간 모델의 그라운드 평면 좌표와 연관된 상기 3D 인간 모델을 해당 이미지 평면 내의 각각의 상기 두 확률 맵들로 다시 투영시킬 수 있다.

[0108] 다른 선택적인 실시예에 있어서, 단일 확률 맵이 다중 카메라들을 위해 이용될 수 있다. 도 17의 예에서, 확률 계산들은 여기서 설명한 바와 같은 각각의 상기 2D 비디오 이미지들에 대해 수행될 수 있고, 각기 상기 해당 2D 이미지 평면에 대응되는 두 이미지 평면 확률 맵들을 생성할 수 있다. 상기 이미지 평면 확률 맵의 확률들은 특정 쓰레시홀드(각 이미지 평면 확률 맵에 대해 동일하거나 다를 수 있는)를 초과하지 않을 경우에 영으로 설정될 수 있다. 각 이미지 평면 확률 맵 내의 식별 좌표들은 각각의 상기 이미지 평면 확률 맵들에 대해 상기 실세계 내의 그라운드 평면 좌표로 해석될 수 있고, 각 비디오 이미지에 대한 그라운드 평면 확률 맵을 생성할 수 있다. 상기 두 이미지 평면 확률 맵들은 병합된 확률 맵을 생성하도록 동일한 그라운드 평면 좌표들을 공유하는 확률들을 곱하여 병합될 수 있다. 상기 병합된 그라운드 평면 확률 맵은 국지적인 최대들을 발견하도록 스캔될 수 있다. 각각의 발견된 국지적인 최대는 이후에 적절하게 $f_1(n)$ 또는 $f_2(n)$(상술한)를 계산하는 데 사용될 수 있는 이들의 해당 이미지 공간 내의 각각의 상기 비디오 이미지들에 대한 별도의 인간 모델들을 확인할 수 있다. 복수의 국지적인 최대들을 위해 상기 병합된 그라운드 평면 확률 맵의 다중 스캔들을 수행하는 것은 후속하는 인간 모델들(각각의 상기 비디오 이미지들에 대한 하나)을 찾고 다음을 계산하도록 수행될 수 있다.

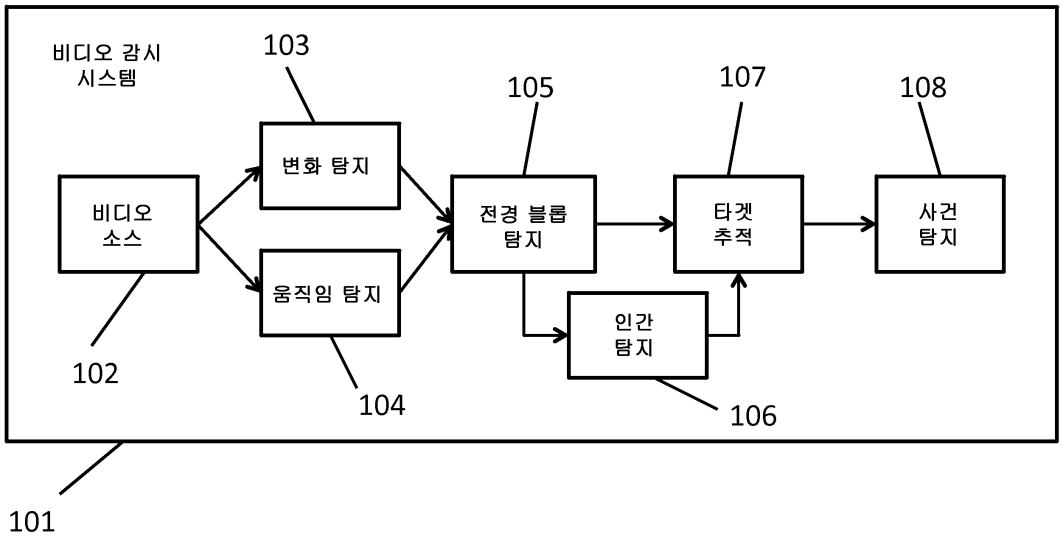
[0109] $$\sum_{n=1}^{m} f_1(n) + f_2(n)$$

[0110] 선택 제한들(최소 확률 쓰레시홀드 및 상기 3D 실세계 내의 최소 거리와 같은)은 변경될 수 있고, 새로운 스캔 패스가 m의 인간 3D 모델들(이러한 예에서 2m의 2D 인간 모델들에 대응되는)의 최적 세트를 찾도록 구현된다.

[0111] 전술한 바에서는 예시적인 실시예들이 설명되었지만 이에 의해 한정되는 것으로 간주되지는 않는다. 비록 몇몇 예시적인 실시예들은 설시하였지만, 해당 기술 분야의 숙련자라면 본 발명의 새로운 교시와 이점들로부터 벗어나지 않고 상기 예시적인 실시예들에서 많은 변형들이 가능한 점을 용이하게 이해할 수 있을 것이다. 예를 들면, 비록 본 발명이 비디오 이미지 내의 인간 객체들의 탐지를 설명하였지만, 본 발명이 이에 한정되는 것으로 간주되어서는 아니 되며, 관심의 대상인 다른 객체들도 탐지될 수 있다.

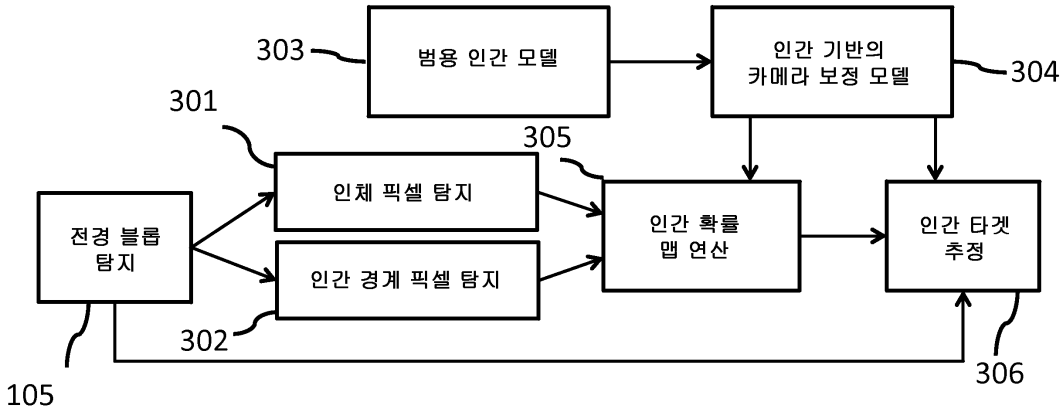
도면

도면1

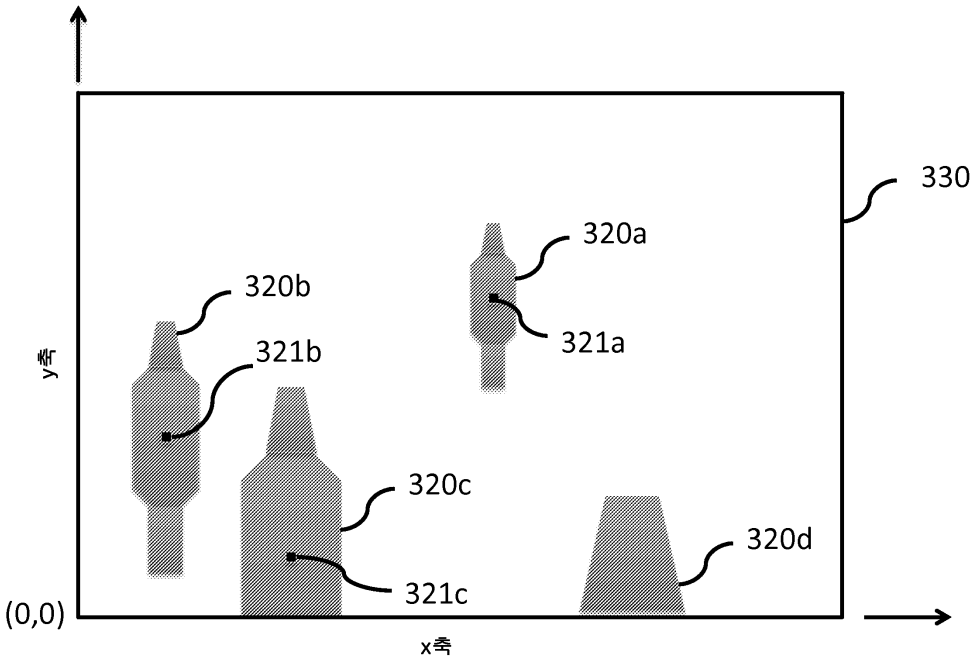


도면2

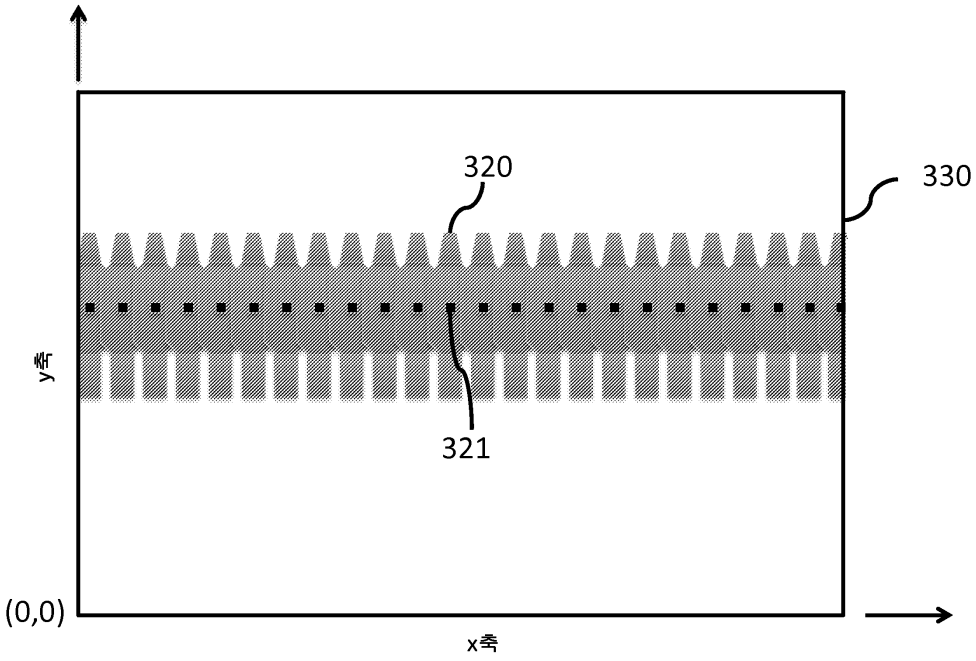
도면3a



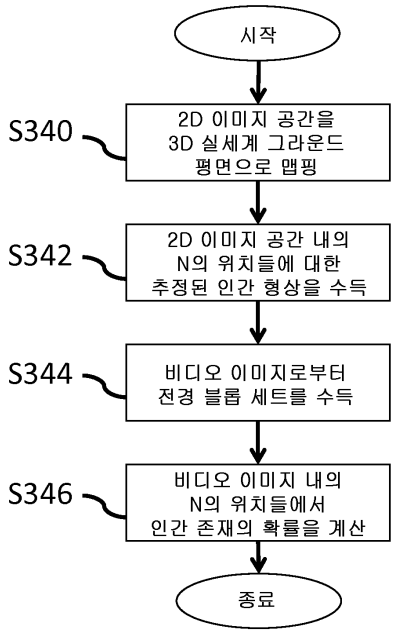
도면3b


330
320a
320b
321a
321b
320c
320d
321c
(0,0)
y축
x축

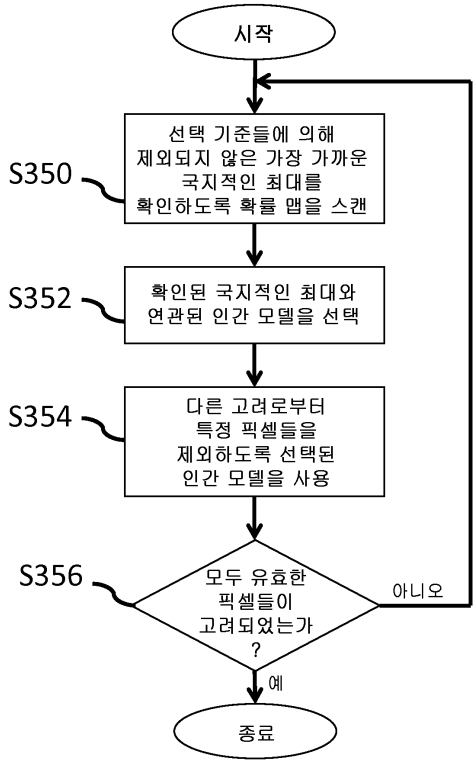
도면3c


320
330
321
(0,0)
y축
x축

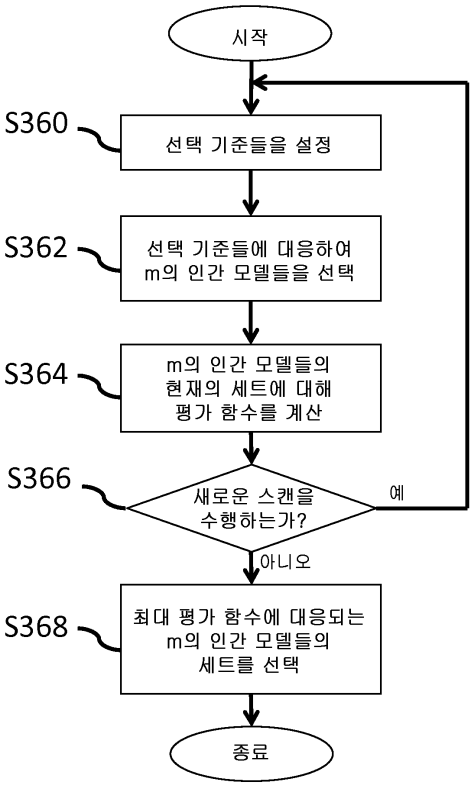
도면3d


시작
S340
2D 이미지 공간을
3D 실세계 그라운드
평면으로 맵핑
S342
2D 이미지 공간 내의
N의 위치들에 대한
추정된 인간 형상을 수득
S344
비디오 이미지로부터
전경 블롭 세트를 수득
S346
비디오 이미지 내의
N의 위치들에서
인간 존재의 확률을 계산
종료

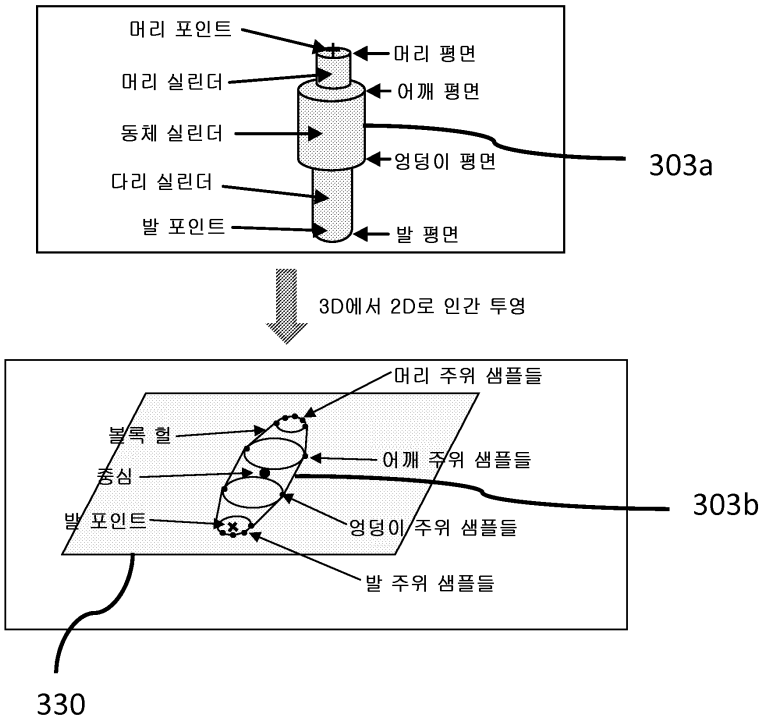
도면3e


시작
S350
선택 기준들에 의해
제외되지 않은 가장 가까운
국지적인 최대를
확인하도록 확률 맵을 스캔
S352
확인된 국지적인 최대와
연관된 인간 모델을 선택
S354
다른 고려로부터
특정 픽셀들을
제외하도록 선택된
인간 모델을 사용
S356
모두 유효한
픽셀들이
고려되었는가
?
아니오
예
종료

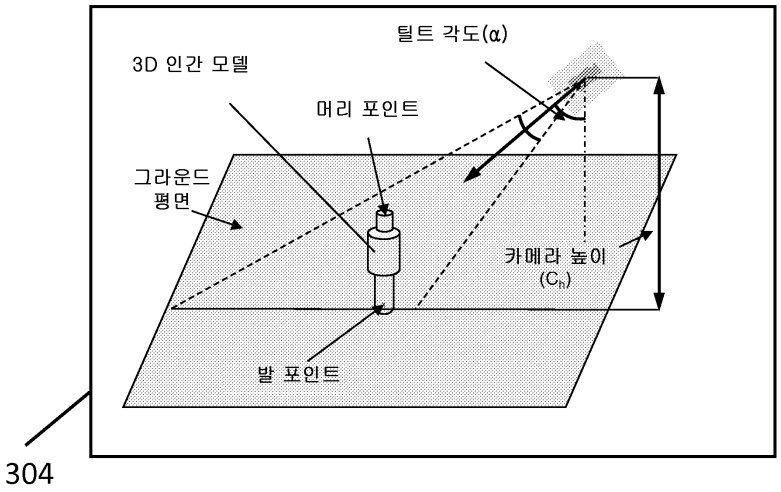
도면3f


시작
S360
선택 기준들을 설정
S362
선택 기준들에 대응하여
m의 인간 모델들을 선택
S364
m의 인간 모델들의
현재의 세트에 대해
평가 함수를 계산
S366
새로운 스캔을
수행하는가?
예
아니오
S368
최대 평가 함수에 대응되는
m의 인간 모델들의
세트를 선택
종료

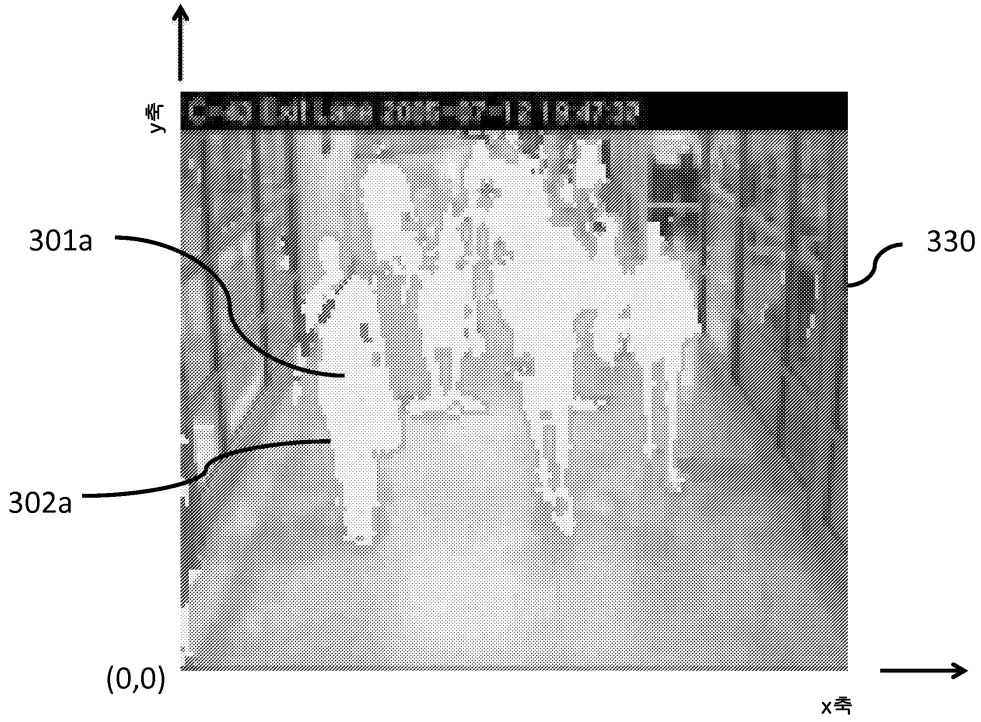
도면4


머리 포인트
머리 실린더
동체 실린더
다리 실린더
발 포인트
머리 평면
어깨 평면
엉덩이 평면
발 평면
303a
3D에서 2D로 인간 투영
머리 주위 샘플들
목 혈
중심
발 포인트
어깨 주위 샘플들
엉덩이 주위 샘플들
발 주위 샘플들
303b
330

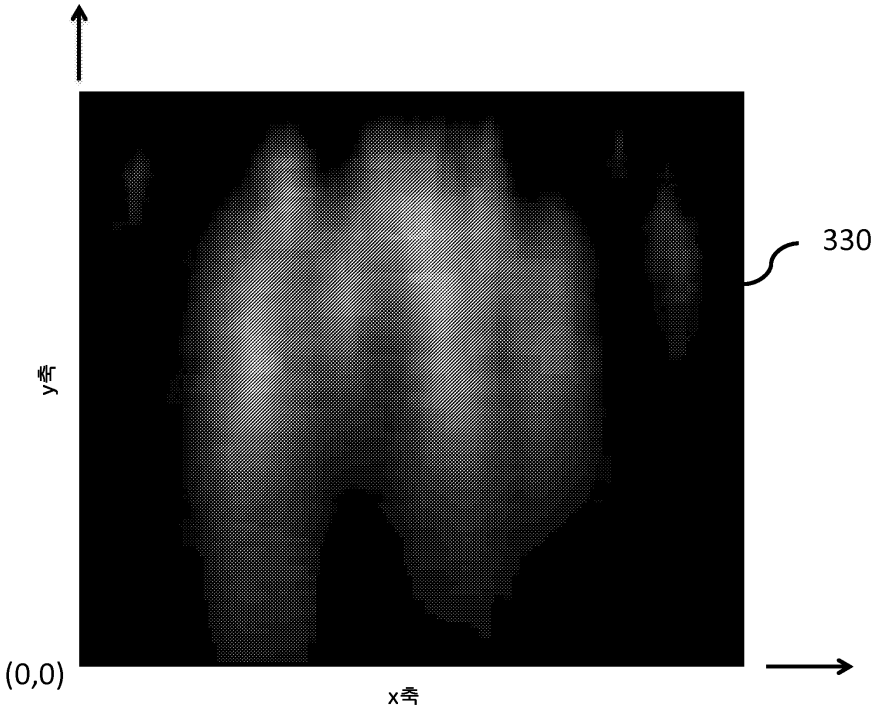
도면5


틸트 각도(α)
3D 인간 모델
머리 포인트
그라운드
평면
카메라 높이
(C_h)
발 포인트
304

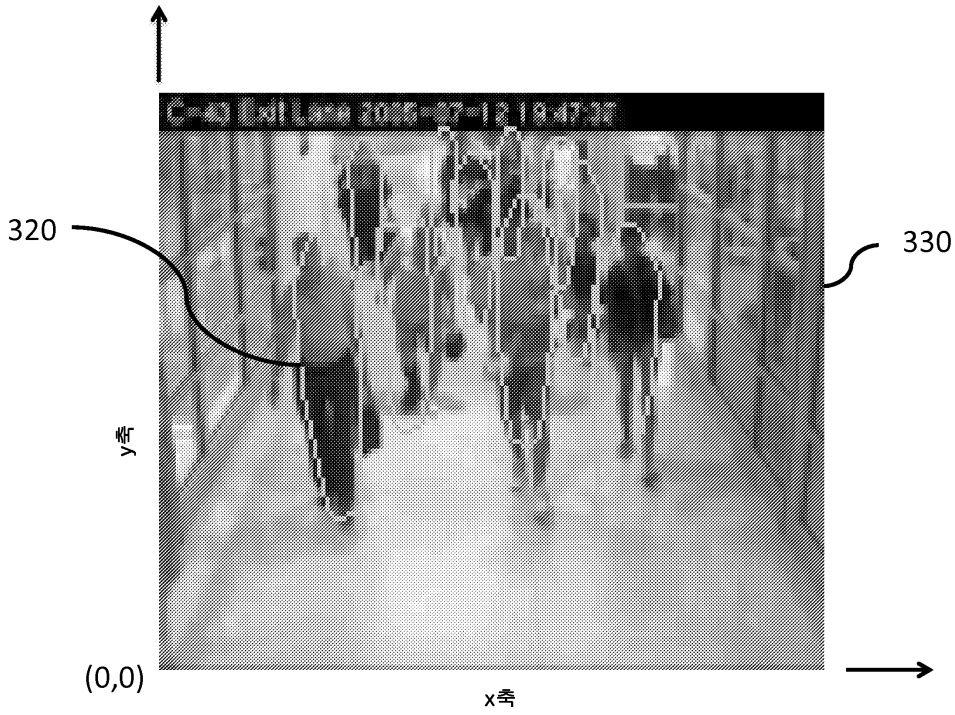
도면6a


y축
301a
330
302a
(0,0)
x축

도면6b

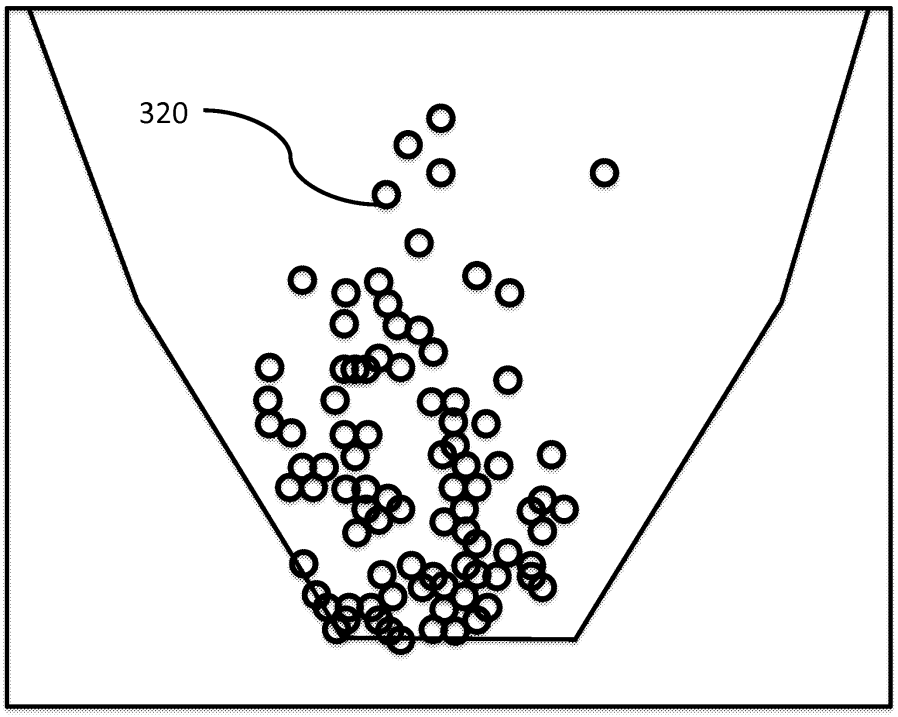


도면6c

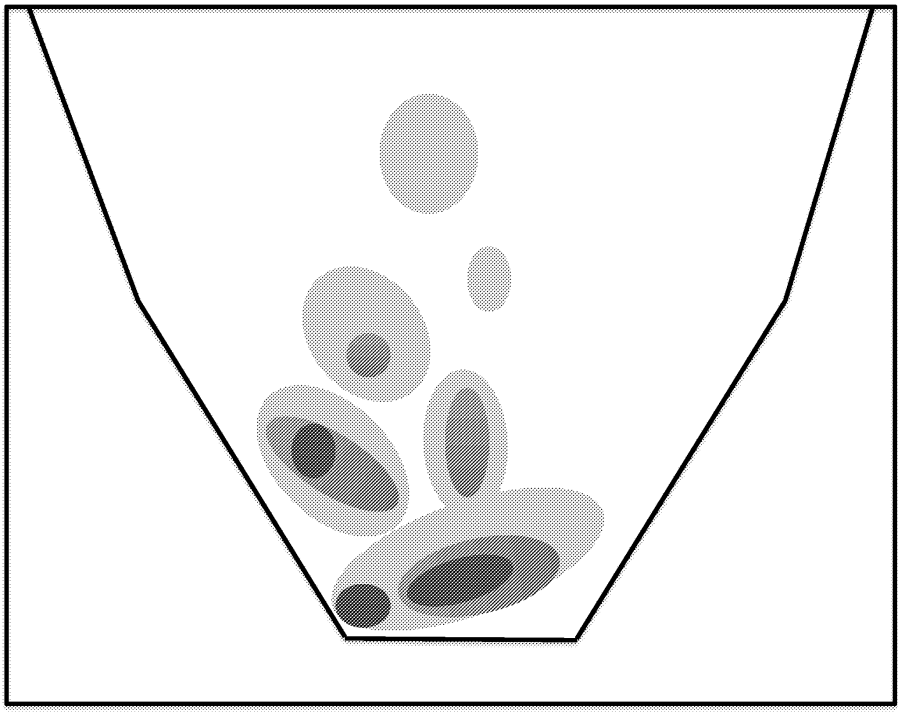


도면7a

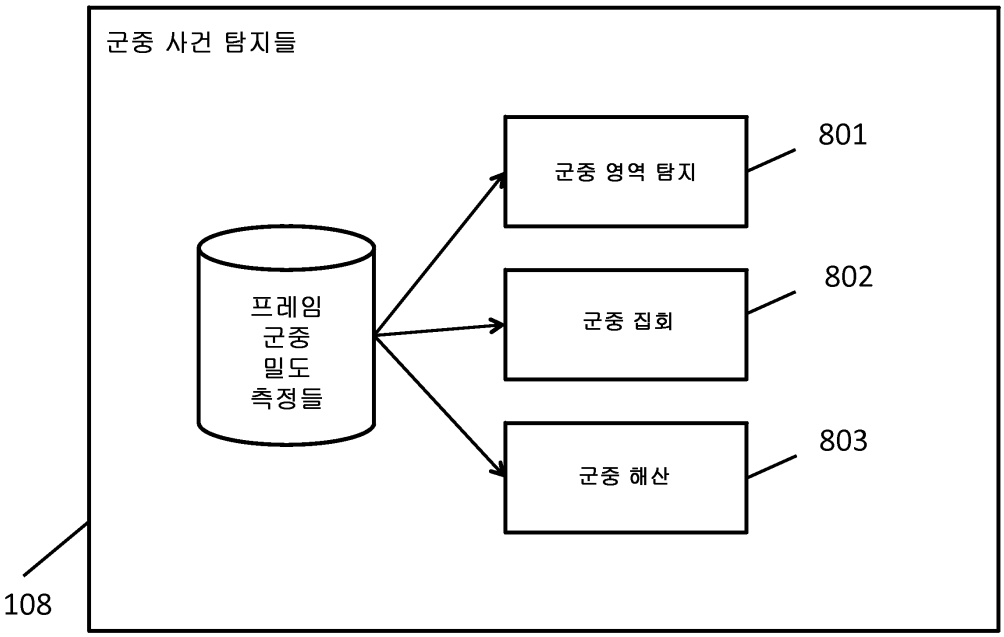
도면7b



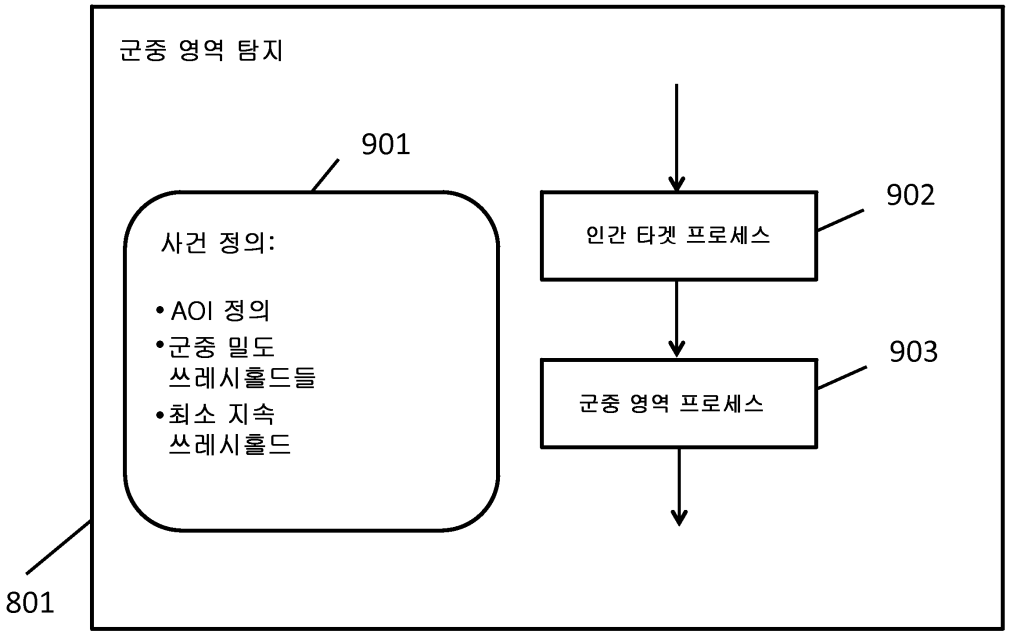
도면7c



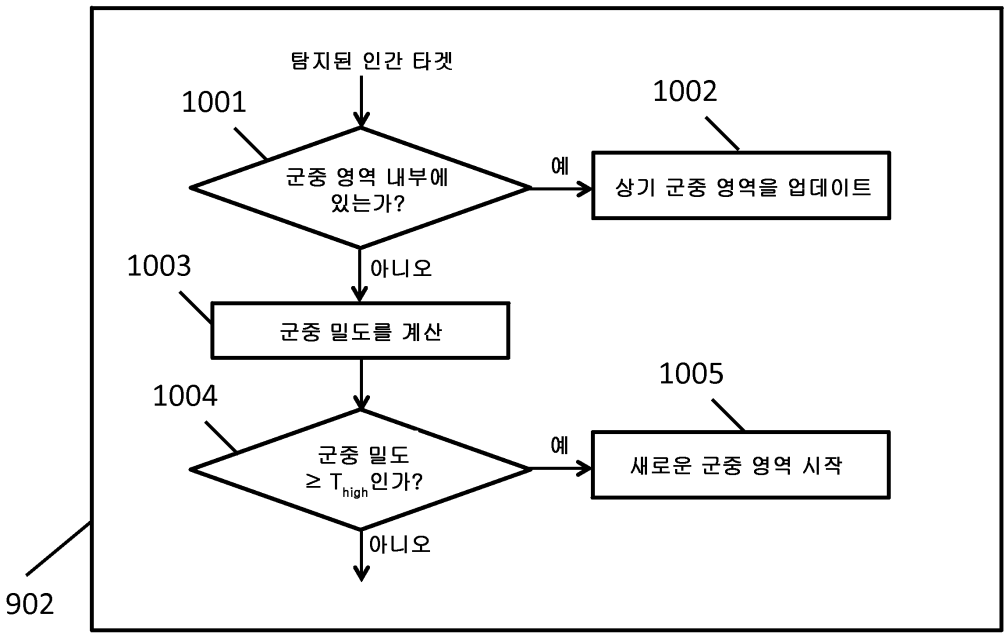
도면8



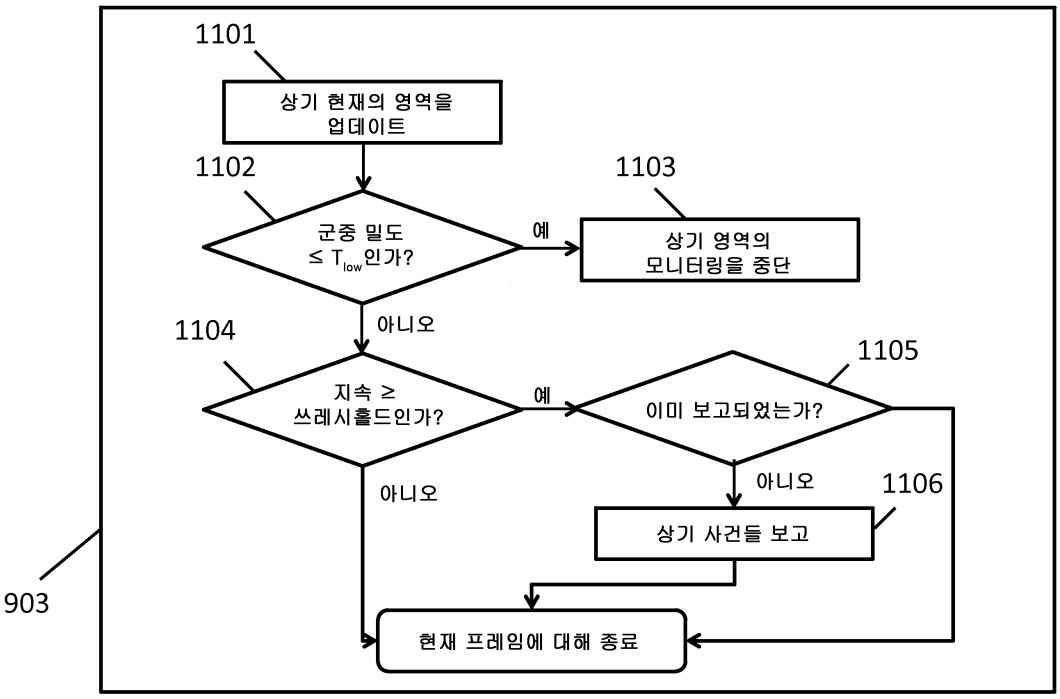
도면9


군중 영역 탐지
901
사건 정의:
• AOI 정의
• 군중 밀도 쓰레시홀드들
• 최소 지속 쓰레시홀드
902
인간 타겟 프로세스
903
군중 영역 프로세스
801

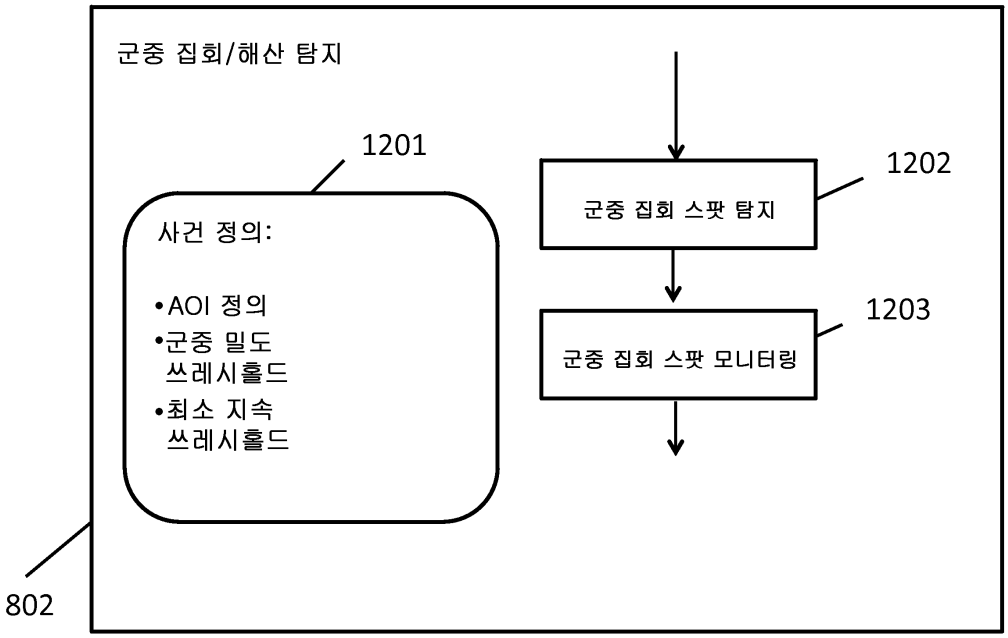
도면10


탐지된 인간 타겟
1001
군중 영역 내부에 있는가?
예
1002
상기 군중 영역을 업데이트
아니오
1003
군중 밀도를 계산
1004
군중 밀도 ≥ T_{high} 인가?
예
1005
새로운 군중 영역 시작
아니오
902

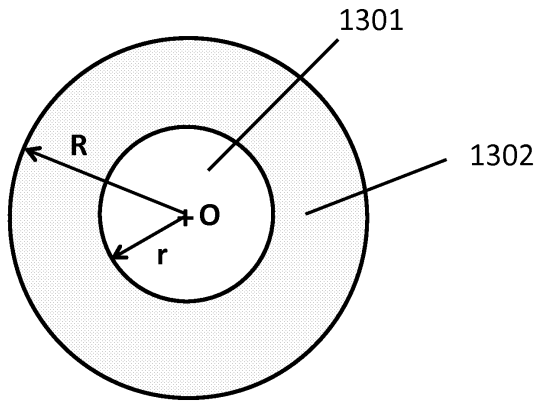
도면11


1101
상기 현재의 영역을 업데이트
1102
군중 밀도 ≤ T_{low}인가?
예
1103
상기 영역의 모니터링을 중단
아니오
1104
지속 ≥ 쓰레시홀드인가?
예
1105
이미 보고되었는가?
아니오
아니오
1106
상기 사건들 보고
현재 프레임에 대해 종료
903

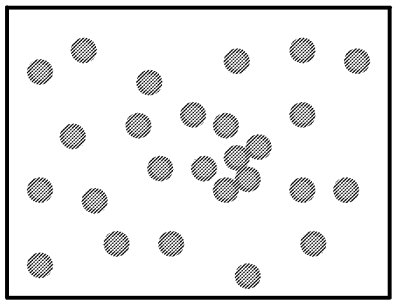
도면12


군중 집회/해산 탐지
1201
사건 정의:
•AOI 정의
•군중 밀도 쓰레시홀드
•최소 지속 쓰레시홀드
1202
군중 집회 스팟 탐지
1203
군중 집회 스팟 모니터링
802

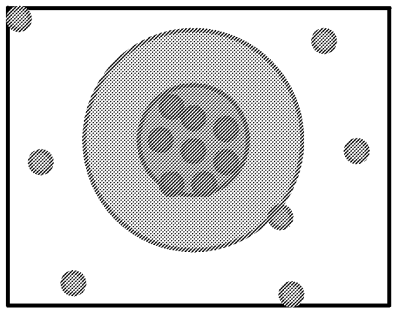
도면13



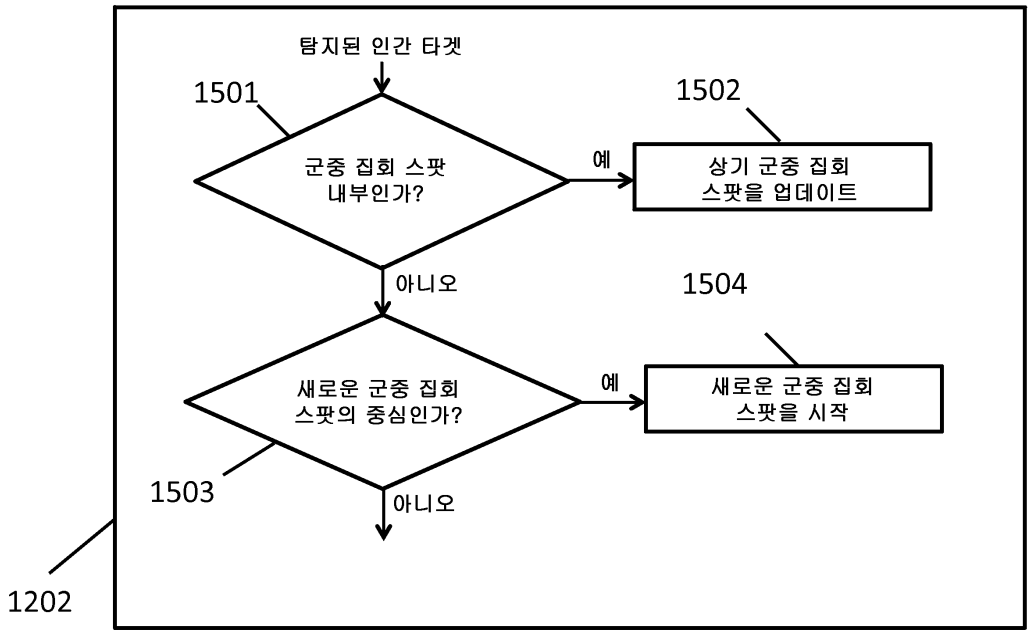
도면14a



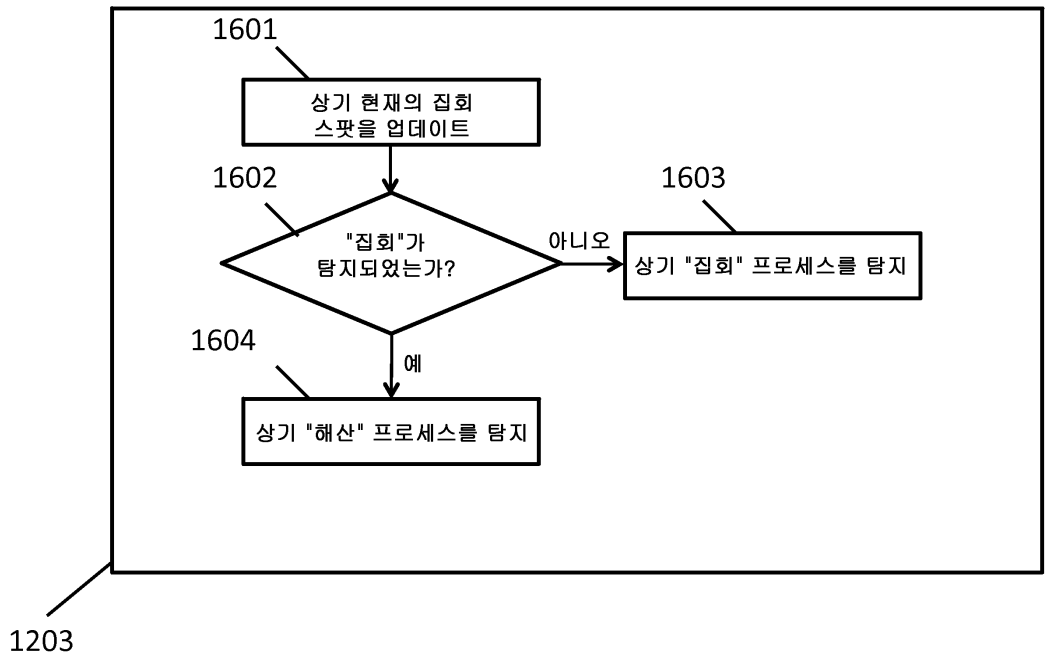
도면14b



도면15


탐지된 인간 타겟
1501
군중 집회 스팟
내부인가?
예
1502
상기 군중 집회
스팟을 업데이트
아니오
1504
새로운 군중 집회
스팟의 중심인가?
예
새로운 군중 집회
스팟을 시작
1503
아니오
1202

도면16


1601
상기 현재의 집회
스팟을 업데이트
1602
"집회"가
탐지되었는가?
아니오
1603
상기 "집회" 프로세스를 탐지
1604
예
상기 "해산" 프로세스를 탐지
1203

도면17