



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2008-0060265
 (43) 공개일자 2008년07월01일

(51) Int. Cl.
 G06K 9/62 (2006.01) G06K 9/46 (2006.01)
 G06T 7/40 (2006.01)
 (21) 출원번호 10-2008-7010536
 (22) 출원일자 2008년04월30일
 심사청구일자 없음
 번역문제출일자 2008년04월30일
 (86) 국제출원번호 PCT/US2006/042062
 국제출원일자 2006년10월27일
 (87) 국제공개번호 WO 2007/053458
 국제공개일자 2007년05월10일
 (30) 우선권주장
 11/263,156 2005년10월31일 미국(US)

(71) 출원인
 이스트맨 코닥 캄파니
 미합중국 뉴욕 로체스터 스테이트 스트리트 343
 (72) 발명자
 갤러거 앤드류 찰스
 미국 뉴욕주 14420 브룩포트 캠프벨 로드 353
 다스 마디라크쉬
 미국 뉴욕주 14612 로체스터 폰드 뷰 하이츠 85
 루이 알렉산더 씨
 미국 뉴욕주 14526 펜필드 세라마르 드라이브 8
 (74) 대리인
 김창세, 김원준

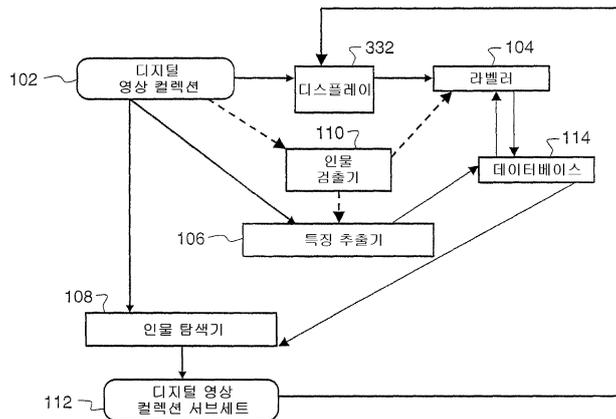
전체 청구항 수 : 총 17 항

(54) 디지털 영상 컬렉션의 특정 인물 식별 방법

(57) 요약

디지털 영상 컬렉션(digital image collection)에서 특정 인물을 식별하는 방법에 관한 것이며, 디지털 영상 컬렉션의 영상 중 하나 이상은 한 명을 초과하는 인물을 포함하며, 이 방법은, 특정 인물 및 하나 이상의 다른 인물을 포함하는 디지털 영상 컬렉션의 제 1 영상에 대하여 하나 이상의 제 1 라벨을 제공하는 단계 - 제 1 라벨은 특정 인물을 식별하고, 특정 인물을 식별하는 디지털 영상 컬렉션의 제 2 영상에 대한 제 2 라벨을 식별함 - 와, 특정 인물을 식별하기 위하여 제 1 라벨과 제 2 라벨을 사용하는 단계와, 제 1 영상 또는 제 2 영상 또는 모두로부터 특정 인물에 관련된 특징을 판정하는 단계와, 특정 인물을 포함한다고 생각되는 디지털 영상의 다른 영상을 식별하기 위하여 이러한 특정 특징을 사용하는 단계를 포함한다.

대표도



특허청구의 범위

청구항 1

디지털 영상 컬렉션(digital image collection)의 특정 인물을 식별하는 방법에 있어서,

상기 디지털 영상 컬렉션의 하나 이상의 영상은 한 명을 초과하는 인물을 포함하고,

상기 방법은,

(a) 특정 인물 및 하나 이상의 다른 인물을 포함하는 상기 디지털 영상 컬렉션의 제 1 영상에 대하여, 하나 이상의 제 1 라벨을 제공하는 단계로서, 상기 제 1 라벨은, 상기 특정 인물을 식별하며, 상기 특정 인물을 식별하는 상기 디지털 영상 컬렉션의 제 2 영상에 대한 제 2 라벨을 식별하는, 상기 하나 이상의 제 1 라벨을 제공하는 단계와,

(b) 상기 특정 인물을 식별하기 위해 상기 제 1 라벨과 제 2 라벨을 사용하는 단계와,

(c) 상기 제 1 영상 또는 제 2 영상 또는 모두로부터 상기 특정 인물에 관련된 특징을 판정하는 단계와,

(d) 상기 특정 인물을 포함한다고 생각되는 상기 디지털 영상 컬렉션의 다른 영상을 식별하기 위하여 이러한 특정 특징을 사용하는 단계를 포함하는

디지털 영상 컬렉션의 특정 인물 식별 방법.

청구항 2

제 1 항에 있어서,

상기 제 1 라벨과 제 2 라벨은 각각 상기 특정 인물의 이름 또는 상기 특정 인물이 상기 제 1 영상과 제 2 영상 안에 모두 있다는 표시를 포함하는

디지털 영상 컬렉션의 특정 인물 식별 방법.

청구항 3

제 1 항에 있어서,

상기 디지털 영상 컬렉션의 상이한 영상에 대응하는 둘보다 많은 라벨이 있는

디지털 영상 컬렉션의 특정 인물 식별 방법.

청구항 4

제 1 항에 있어서,

사용자가 상기 제 1 라벨과 제 2 라벨을 제공하는

디지털 영상 컬렉션의 특정 인물 식별 방법.

청구항 5

제 1 항에 있어서,

단계 (c)는 상기 특정 인물의 상기 특징을 판정하기 위하여 상기 영상 안의 인물을 검출하는 단계를 포함하는

디지털 영상 컬렉션의 특정 인물 식별 방법.

청구항 6

제 4 항에 있어서,

영상 안의 상기 특정 인물의 위치는 상기 사용자에게 의해 제공되지 않는

디지털 영상 컬렉션의 특정 인물 식별 방법.

청구항 7

제 4 항에 있어서,

상기 디지털 영상 컬렉션의 하나 이상의 상기 영상 안의 상기 특정 인물의 위치는 상기 사용자에게 의해 제공되는 디지털 영상 컬렉션의 특정 인물 식별 방법.

청구항 8

제 1 항에 있어서,

상기 제 1 라벨은 상기 특정 인물의 이름 및 상기 제 1 영상 안의 특정 인물의 위치를 포함하고,

상기 제 2 라벨은 다수의 인물을 포함하는 상기 제 2 영상 안에 상기 특정 인물이 있다는 것을 표시하는 디지털 영상 컬렉션의 특정 인물 식별 방법.

청구항 9

제 8 항에 있어서,

다수의 상이한 인물을 식별하기 위한 다수의 라벨이 있는

디지털 영상 컬렉션의 특정 인물 식별 방법.

청구항 10

제 9 항에 있어서,

특정 인물 및 영상 안의 그 인물의 위치를 식별하는 라벨을 사용자가 제공하며, 상기 다수의 라벨이 상기 특정 인물을 포함하는 이 영상을 식별하기 위해 사용되며, 상기 특징을 판정하기 위하여 상기 사용된 식별된 인물을 분석하는

디지털 영상 컬렉션의 특정 인물 식별 방법.

청구항 11

제 10 항에 있어서,

각 라벨은 상기 특정 인물의 이름을 포함하는

디지털 영상 컬렉션의 특정 인물 식별 방법.

청구항 12

제 1 항에 있어서,

(e) 상기 사용자에게 상기 특정 인물을 포함한다고 생각되는 영상을 디스플레이하는 단계와,

(f) 상기 특정 인물이 상기 디스플레이된 영상에 포함되어 있는지 확인하기 위해 상기 사용자가 상기 디스플레이된 영상을 검토하는 단계를 더 포함하는

디지털 영상 컬렉션의 특정 인물 식별 방법.

청구항 13

디지털 영상 컬렉션의 특정 인물을 식별하는 방법에 있어서,

상기 디지털 영상 컬렉션의 하나 이상의 영상은 한 명을 초과하는 인물을 포함하고,

상기 방법은,

(a) 특정 인물을 포함하는 영상에 대하여 하나 이상의 라벨을 제공하는 단계로서, 상기 라벨은 상기 특정 인물을 포함하는 상기 영상을 식별하는, 상기 하나 이상의 라벨을 제공하는 단계와,

- (b) 상기 특정 인물에 관련된 특징을 판정하는 단계와,
 - (c) 상기 특정 인물을 포함하는 것으로 생각되는 상기 컬렉션의 영상을 식별하기 위하여 이러한 특정 인물 특징 및 상기 라벨을 사용하는 단계와,
 - (d) 상기 사용자에게 상기 특정 인물을 포함하는 것으로 생각되는 영상을 디스플레이하는 단계와,
 - (e) 상기 특정 인물이 상기 디스플레이된 영상에 포함되어 있는지 확인하기 위해 상기 사용자가 상기 디스플레이된 영상을 검토하는 단계를 포함하는
- 디지털 영상 컬렉션의 특정 인물 식별 방법.

청구항 14

제 13 항에 있어서,
상기 사용자가 상기 특정 인물이 상기 디스플레이된 영상에 포함되어 있다는 것을 확인하였을 때, 상기 사용자가 라벨을 제공하는
디지털 영상 컬렉션의 특정 인물 식별 방법.

청구항 15

제 14 항에 있어서,
상기 판정된 특징은 상기 사용자가 제공한 라벨을 사용하여 업데이트되는
디지털 영상 컬렉션의 특정 인물 식별 방법.

청구항 16

제 1 항에 있어서,
상기 특징은 얼굴 측정치, 옷 또는 안경 또는 그것으로부터의 조합으로부터 판정되는
디지털 영상 컬렉션의 특정 인물 식별 방법.

청구항 17

제 13 항에 있어서,
상기 특징은 얼굴 측정치, 옷 또는 안경 또는 그것으로부터의 조합으로부터 판정되는
디지털 영상 컬렉션의 특정 인물 식별 방법.

명세서

기술분야

- <1> 본 발명은 일반적으로 영상 처리 분야에 관한 것이다. 보다 구체적으로는, 본 발명은 포착된 영상의 그 소실점 (vanishing points)의 대응하는 위치에 기초하여, 영상 포착 때 일어나는 의도하지 않은 회전의 카메라 앵글에 대한 추정 및 보정에 관련된다. 더욱이, 본 발명은 디지털 카메라에서 이러한 영상 처리를 수행하는 것에 관한 것이다.
- <2> 본 발명은 만약 관심이 있는 물체나 인물이 디지털 영상의 컬렉션(collection)의 특정 영상 안에 있는지 판정하는 것에 관한 것이다.

배경기술

- <3> 디지털 사진술의 출현과 함께, 소비자는 디지털 영상 및 비디오의 커다란 컬렉션을 모으고 있다. 사진사당 디지털 카메라를 사용한 영상 포착의 평균 수는 여전히 매년 증가하고 있다. 그 결과, 영상 및 비디오의 구성 (organization) 및 검색(retrieval)은 이미 통상의 소비자에 대한 문제가 되었다. 현재, 통상의 소비자의 디지털 영상 컬렉션이 걸치는 시간의 길이는 단지 몇 년밖에 되지 않는다. 구성 및 검색 문제는 평균 디지털 영상

및 비디오 컬렉션이 걸치는 시간의 길이가 증가할수록 계속 커질 것이다.

- <4> 사용자는 관심이 있는 특정 인물을 포함하는 영상 및 비디오를 탐색하는 것을 원한다. 사용자는 관심이 있는 인물을 포함하는 영상 및 비디오를 찾기 위하여 수동 검색(a manual search)을 수행할 수 있다. 그러나 이는 느리고 힘든 프로세스이다. 비록 몇몇 상업적 소프트웨어(예, 아도브 앨범(Adobe Album))가 나중에 검색될 수 있도록 영상 안의 인물을 표시하는 라벨을 영상에 다는 것을 허용하지만, 초기의 라벨을 붙이는 프로세스(labeling process)는 여전히 매우 지루하며 시간을 소비한다.
- <5> 얼굴 인식 소프트웨어는 그라운드-트루쓰(ground-truth) 라벨 붙여진 영상의 세트(즉, 대응하는 인물 신원(identity)이 있는 영상의 세트)의 존재를 가정한다. 대다수의 소비자 영상 컬렉션은 유사한 그라운드 트루쓰(ground truth)를 구비하지 않는다. 또한, 영상 안의 얼굴을 라벨 붙이는 것은 많은 소비자 영상이 다수의 인물을 가지고 있으므로 복잡하다. 따라서 단순히 영상을 영상 안의 인물의 신원으로 라벨 붙이는 것은 영상 안의 어떤 인물이 어떤 신원에 관련되는지 표시하지 않는다.
- <6> 보안 또는 다른 목적으로 인물을 인식하기 위해 시도하는 많은 영상 처리 패키지가 존재한다. 몇몇 예는 코니텍 시스템즈 게엠베하(Conitec Systems GmbH)의 FaceVACS 얼굴 인식 소프트웨어 및 이미지스 테크놀로지 인크(Imagis Technologies Inc.) 및 아이덴티스 인크(Identix Inc.)의 얼굴 인식 SDK(Facial Recognition SDKs)이다. 이러한 패키지는 주로 인물이 균일한 조명(uniform illumination), 정면 자세(frontal pose) 및 특성 없는 표정(neutral expression) 아래에서 카메라를 직시하는 보안-유형(security-type) 애플리케이션 목적에 사용된다. 이러한 방법은 자세(pose), 조명(illumination), 표정(expression) 및 얼굴 크기(face size)에 있어서의 큰 변화가 이 범위(domain) 안의 영상에 직면하게 되므로 개인적인 소비자 영상에의 사용에는 적합하지 않다.

발명의 상세한 설명

- <7> 본 발명의 목적은 디지털 영상 컬렉션의 영상 또는 비디오 안에서 관심이 있는 물체 또는 인물을 쉽게 식별하기 위한 것이다. 이 목적은 디지털 영상 컬렉션의 특정 인물을 식별하는 방법에 의해 달성되며, 디지털 영상 컬렉션의 하나 이상의 영상은 한 명을 초과하는 인물을 포함하며,
- <8> 이 방법은,
- <9> (a) 특정 인물 및 하나 이상의 다른 인물을 포함하는 디지털 영상 컬렉션의 제 1 영상에 대하여 하나 이상의 제 1 라벨을 제공하는 단계 - 제 1 라벨은 특정 인물을 식별하고, 특정 인물을 식별하는 디지털 영상 컬렉션의 제 2 영상에 대한 제 2 라벨을 식별함 - 와,
- <10> (b) 특정 인물을 식별하기 위해 제 1 라벨 및 제 2 라벨을 사용하는 단계와,
- <11> (c) 제 1 영상 또는 제 2 영상 또는 모두로부터 특정 인물에 관련되는 특징을 판정하는 단계와,
- <12> (d) 특정 인물을 포함한다고 생각되는 디지털 영상 컬렉션의 다른 영상을 식별하기 위하여 이러한 특정 특징을 사용하는 단계
- <13> 를 포함한다.
- <14> 이 방법은 사용자들이 관심이 있는 인물을 사용하기 쉬운 인터페이스로 탐색하는 것을 허용하는 장점을 가진다. 더욱이, 이 방법은 영상이 자동으로 관심이 있는 인물에 관련된 라벨로 라벨 붙여지고, 사용자들이 라벨을 재검토하는 것을 허용하는 장점을 가진다.

실시예

- <36> 다음 설명에서, 본 발명의 몇몇 실시예가 소프트웨어 프로그램으로 설명될 것이다. 당업자는 이와 같은 방법의 균등물이 본 발명의 범주 안에서 하드웨어 또는 소프트웨어로 또한 구성될 수 있음을 쉽게 인식할 것이다.
- <37> 영상 조작(image manipulation) 알고리즘 및 시스템이 잘 알려져 있으므로, 본 설명은 본 발명에 따르는 방법의 일부를 구성하거나, 또는 더 직접적으로 협력하는 알고리즘 및 시스템에 특히 기울여질 것이다. 여기에 구체적으로 도시되거나 설명되지 않은, 이러한 알고리즘 및 시스템의 다른 특징 및 그것과 함께 포함된 영상 신호를 생성하거나 달리 처리하기 위한 하드웨어 또는 소프트웨어는 해당 기술에 알려진 이러한 시스템, 알고리즘, 구성 요소(components) 및 요소(elements)로부터 선택될 수 있다. 다음 명세서에 설명된 것과 같은 설명이 주어진다. 그것으로부터의 모든 소프트웨어 구현은 통상적이며 이러한 해당 기술의 통상의 기술 안에 있다.

- <38> 도 1은 본 발명을 구현할 수 있는 디지털 카메라 폰(301) 기반 영상 시스템의 블록도이다. 디지털 카메라 폰(301)은 디지털 카메라의 하나의 유형이다. 바람직하게는, 디지털 카메라 폰(301)은 충분히 작아 영상을 포착하고 재검토할 때 사용자에게 의해 쉽게 손에 잡히는, 휴대용 배터리에 의해 동작하는 디바이스이다. 디지털 카메라 폰(301)은 영상 데이터/메모리(330)를 사용하여 저장된 디지털 영상을 생성하며, 이는 예를 들어, 내부 플래시 EPROM 메모리(internal Flash EPROM memory) 또는 착탈 가능 메모리 카드(a removable memory card)일 수 있다. 자기 하드 드라이브(magnetic hard drives), 자기 테이프(magnetic tape) 또는 광 디스크(optical disks)와 같은 다른 유형의 디지털 영상 저장 매체가 영상/데이터 메모리(330)를 제공하기 위해 대신 사용될 수 있다.
- <39> 디지털 카메라 폰(301)은 스크린(도시하지 않음)으로부터 CMOS 영상 센서(311)의 영상 센서 어레이(314) 위로 빛을 초점에 모으는 렌즈(305)를 포함한다. 영상 센서 어레이(314)는 잘 알려진 베이어 컬러 필터 패턴(Bayer color filter pattern)을 사용하여 컬러 영상 정보를 제공할 수 있다. 타이밍 생성기(312)는 영상 센서 어레이(314)를 제어하며, 이는 또한 주위의 조명이 낮을 때 스크린을 밝게 하기 위하여 플래시(303)를 제어한다. 영상 센서 어레이(314)는 예를 들어, 1280 열(columns) x 960 행(rows)의 픽셀을 가질 수 있다.
- <40> 몇몇 실시예에서, 디지털 카메라 폰(301)은 또한 저 해상도의 비디오 영상 프레임을 생성하기 위해 영상 센서 어레이(314)의 다수의 픽셀을 함께 합함으로써(예, 영상 센서 어레이(314)의 각 4 열 x 4 행 영역 안의 동일한 컬러의 픽셀을 합함), 비디오 클립을 저장할 수 있다. 비디오 영상 프레임은 예를 들어, 초당 24 프레임의 판독률(readout rate)을 사용하여, 정규 간격에서 영상 센서 어레이(314)로부터 읽혀질 수 있다.
- <41> 영상 센서 어레이(314)로부터의 아날로그 출력 신호는 CMOS 영상 센서(311) 위의 아날로그-투-디지털(A/D) 컨버터 회로(316)에 의해 증폭되고 디지털 데이터로 변환된다. 디지털 데이터는 DRAM 버퍼 메모리(318)에 저장되고 플래시 EPROM 메모리 일 수 있는 펌웨어 메모리(328)에 저장된 펌웨어에 의해 제어되는 디지털 프로세서(320)에 의해 순차적으로 처리된다. 디지털 프로세서(320)는 디지털 카메라 폰(301) 및 디지털 프로세서(320)가 저 전력 상태에 있을 때에도 날짜 및 시간을 유지하는, 리얼-타임 클락(real-time clock)(324)을 포함한다.
- <42> 처리된 디지털 영상 파일은 영상/데이터 메모리(330) 안에 저장된다. 영상/데이터 메모리(330)는 또한 도 11을 참조하여 추후에 설명될 것과 같이, 사용자의 개인적인 캘린더 정보를 저장하기 위하여 사용될 수 있다. 영상/데이터 메모리는 또한 폰 번호, 해야 할 일 목록(to-do lists) 및 이와 유사한 것과 같은 다른 유형의 데이터를 저장할 수 있다.
- <43> 스틸 영상 모드(still image mode)에서, 디지털 프로세서(320)는 연출된(rendered) sRGB 영상 데이터를 생성하기 위해, 컬러 및 톤 보정이 뒤따르는 컬러 인터폴레이션(color interpolation)을 수행한다. 디지털 프로세서(320)는 또한 사용자에게 의해 선택된 다양한 영상 크기를 제공할 수 있다. 연출된 sRGB 영상 데이터는 그리고 나서 JPEG 압축되어 영상/데이터 메모리(330)에 JPEG 영상 파일로 저장된다. JPEG 파일은 미리 설명된 소위 "Exif" 영상 포맷을 사용한다. 이 포맷은 다양한 TIFF 태그를 사용하여 특정 영상 메타데이터(metadata)를 저장하는 Exif 애플리케이션 세그먼트(segment)를 포함한다. 별도의 TIFF 태그는 예를 들어, 사진이 포착된 날짜 및 시간, 렌즈 f/수(f/number) 및 다른 카메라 설정을 저장하고, 영상 캡션(caption)을 저장하기 위하여 사용될 수 있다. 특히, 영상설명(ImageDescription) 태그는 라벨을 저장하기 위하여 사용될 수 있다. 리얼-타임 클락(real-time clock)(324)은 포착 날짜/시간 값을 제공하며, 이는 각 Exif 영상 파일에 날짜/시간 메타데이터로 저장된다.
- <44> 위치 판정기(a location determiner)(325)는 영상 포착과 관련된 위치 정보를 제공한다. 위치는 바람직하게는 위도 및 경도 단위로 저장된다. 위치 판정기(325)는 영상 포착 시간과 조금 다른 시간에 지리적 위치를 판정할 수 있다는 것을 유의하라. 이러한 경우에서, 위치 판정기(325)는 가장 가까운 시간으로부터의 지리적 위치를 영상과 관련된 지리적 위치로 사용할 수 있다. 이와 달리, 위치 판정기(325)는 영상 포착과 관련된 지리적 위치를 판정하기 위하여 영상 포착 시간 전 및/또는 후의 시간의 다수의 지리적 위치 사이에서 인터폴레이트(interpolate) 할 수 있다. 인터폴레이션(interpolation)은 위치 판정기(325)가 항상 지리적 위치를 판정하는 것이 가능하지는 않기 때문에 필요로 될 수 있다. 예를 들어, 실내에 있을 때, GPS 수신기는 종종 신호를 검출하는데 실패한다. 이러한 경우에서, 최근의 성공적인 지리적 위치(즉, 건물 안에 들어가기 전)가 특정 영상 포착과 관련된 지리적 위치를 추정하기 위하여 위치 판정기(325)에 의해 사용될 수 있다. 위치 판정기(325)는 영상의 위치를 판정하기 위한 수많은 방법 중 어떠한 것이라도 사용할 수 있다. 예를 들어, 지리적 위치는 잘-알려진(well-known) 전 지구 위치 확인 위성(Global Positioning Satellites(GPS))으로부터 통신을 수신함으로써 판정될 수 있다.

- <45> 디지털 프로세서(320)는 또한 저-해상도 "섬네일(thumbnail)" 크기 영상을 생성하며, 이는 그 명세서가 여기에 참조로 인용되어 있고, 쿠츠타(Kuchta) 등에게 함께-양도된(commonly-assigned) 미국 특허 제 5,164,831호에 설명된 것과 같이 생성될 수 있다. 섬네일(thumbnail) 영상은 RAM 메모리(322)에 저장되어 예를 들어, 액티브 매트릭스 LCD(an active matrix LCD) 또는 유기 발광 다이오드(organic light emitting diode(OLED)) 일 수 있는 컬러 디스플레이(322)에 공급될 수 있다. 영상이 포착된 후, 이들은 섬네일 영상 데이터를 사용하여 컬러 LCD 영상 디스플레이(322) 위에서 빨리 재검토될 수 있다.
- <46> 컬러 디스플레이(322) 위에 디스플레이되는 그래픽 사용자 인터페이스(graphical user interface)는 사용자 컨트롤(334)에 의해 제어된다. 사용자 컨트롤(334)은 폰 번호에 전화를 걸기 위한 전용의 푸쉬 버튼(예, 전화 키 패드), 모드(예, "폰(phone)" 모드, "카메라(camera)" 모드)를 설정하기 위한 컨트롤, 4 방향 컨트롤(위, 아래, 좌, 우)을 포함하는 조이스틱 컨트롤러 및 푸시-버튼 가운데의 "오케이(OK)" 스위치, 또는 이와 유사한 것을 포함할 수 있다.
- <47> 디지털 프로세서(320)에 연결된 오디오 코덱(340)은 마이크로폰(342)으로부터 오디오 신호를 수신하고 스피커(344)로 오디오 신호를 제공한다. 이 구성 요소는 전화 대화를 위한 것 및 비디오 시퀀스(video sequence) 또는 스틸 영상(still image)과 함께, 오디오 트랙을 재생하고 저장하기 위한 것에서 모두 사용될 수 있다. 스피커(344)는 또한 사용자에게 들어오는 폰 콜(incoming phone call)을 알려주기 위해 사용될 수 있다. 이는 펌웨어 메모리(328)에 저장된 표준 링 톤(a standard ring tone)을 사용하여 또는 모바일 폰 네트워크(358)로부터 다운로드되고 영상/데이터 메모리(330)에 저장된 커스텀 링-톤(a custom ring-tone)을 사용하여 행해질 수 있다. 추가로, 진동 디바이스(a vibration device)(도시하지 않음)가 들어오는 폰 콜(incoming phone call)의 조용한(예, 들리지 않게) 통지를 제공하기 위하여 사용될 수 있다.
- <48> 독 인터페이스(a dock interface)(362)는 디지털 카메라 폰(301)을 일반 제어 컴퓨터(a general control computer)(40)에 연결되어 있는 독/충전기(a dock/charger)(364)에 연결하기 위하여 사용될 수 있다. 독 인터페이스(362)는 예를 들어, 잘-알려진 USB 인터페이스 사양을 따를 수 있다. 이와 달리, 디지털 카메라 폰(301) 및 일반 제어 컴퓨터(40) 사이의 인터페이스는 잘-알려진 블루투스(Bluetooth) 무선 인터페이스 또는 잘-알려진 802.11b 무선 인터페이스와 같은 무선 인터페이스 일 수 있다. 독 인터페이스(362)는 영상/데이터 메모리(330)로부터 일반 제어 컴퓨터(40)로 영상을 다운로드하기 위하여 사용될 수 있다. 독 인터페이스(362)는 또한 일반 제어 컴퓨터(40)로부터 디지털 카메라 폰(301) 안의 영상/데이터 메모리로 캘린더 정보를 다운로드하기 위하여 사용될 수 있다. 독/충전기(364)는 또한 디지털 카메라 폰(301)의 배터리(도시하지 않음)를 재충전하기 위해서 사용될 수 있다.
- <49> 디지털 프로세서(320)는 무선 모뎀(350)에 연결되며 이는 디지털 카메라 폰(301)이 RF 채널(352)을 통해 정보를 전송하고 수신하는 것을 가능하게 한다. 무선 모뎀(350)은 3GSM 네트워크와 같은 모바일 폰 네트워크(358)와 라디오 주파수(예, 무선) 링크를 통해 통신한다. 모바일 폰 네트워크(358)는 디지털 카메라 폰(301)으로부터 업로드된 디지털 영상을 저장할 수 있는 포토 서비스 제공기(a photo service provider)(372)와 통신한다. 이 영상은 일반 제어 컴퓨터(40)를 포함하는 다른 디바이스에 의해 인터넷(370)을 통해 액세스 될 수 있다. 모바일 폰 네트워크(358)는 또한 통상적인 전화 서비스를 제공하기 위하여 표준 전화 네트워크(a standard telephone network)(도시하지 않음)에 연결된다.
- <50> 본 발명의 일 실시예가 도 2에 도시되어 있다. 인물을 포함하는 디지털 영상 컬렉션(a digital image collection)(102)이 인물 탐색기(a person finder)(108)에 의해 관심이 있는 인물에 대하여 검색된다. 디지털 영상 컬렉션 서브세트(a digital image collection subset)(112)는 관심이 있는 인물을 포함한다고 생각되는 디지털 영상 컬렉션(102)로부터의 영상의 세트(set)이다. 디지털 영상 컬렉션(102)는 디지털 및 비디오 모두를 포함한다. 편의상, "영상(image)"이라는 용어는 단일 영상 및 비디오 모두를 지칭한다. 비디오는 오디오 또는 때로는 텍스트를 수반하는 영상의 컬렉션이다. 디지털 영상 컬렉션 서브세트(112)는 인간 사용자에 의해 재검토되기 위해 디스플레이(332) 위에 디스플레이된다.
- <51> 관심이 있는 인물에 대한 검색은 사용자에 의해 다음과 같이 개시된다. 디지털 영상 컬렉션(102)의 영상 또는 비디오는 디스플레이(332) 위에 디스플레이되고 사용자에 의해 검토된다. 사용자는 라벨러(a labeler)(104)로 하나 이상의 영상에 대해 하나 이상의 라벨을 수립한다. 특징 추출기(a feature extractor)(106)는 라벨러(104)로부터의 라벨과 관련하여 디지털 영상 컬렉션으로부터 특징을 추출한다. 특징은 라벨과 관련하여 데이터베이스(114)에 저장된다. 인물 검출기(a person detector)(110)는 라벨 붙이기(labeling) 및 특징 추출(feature extraction)을 돕기 위하여 선택적으로 사용될 수 있다. 디지털 영상 컬렉션 서브세트(112)가 디스플레이(332)

위에 디스플레이 될 때, 사용자는 결과를 재검토할 수 있으며 디스플레이된 영상에 더 라벨 붙일 수 있다.

- <52> 라벨러(104)로부터의 라벨은 특정 영상 또는 비디오가 관심이 있는 인물을 포함한다는 것을 표시하며 다음 중 하나 이상을 포함한다.
- <53> (1) 영상 또는 비디오 안의 관심이 있는 인물의 이름. 인물의 이름은 이름(a given name) 또는 별명(a nickname)일 수 있다.
- <54> (2) "인물 A(Person A)" 또는 "인물 B(Person B)"와 같은 식별자(identifier) 또는 텍스트 열(text string)과 같은 관심이 있는 인물과 관련되는 식별자(an identifier).
- <55> (3) 영상 또는 비디오 안의 관심이 있는 인물의 위치. 바람직하게는, 관심이 있는 인물의 위치는 관심이 있는 인물의 눈의 좌표(예, 행 및 열의 픽셀 주소)(및 비디오의 경우 관련된 프레임 번호)에 의해 특정된다. 이와 달리, 관심이 있는 인물의 위치는 관심이 있는 인물의 몸 또는 얼굴을 둘러싸는 박스의 좌표에 의해 특정될 수 있다. 다른 대안으로, 관심이 있는 인물의 위치는 관심이 있는 인물의 안에 포함된 위치를 표시하는 좌표에 의해 특정될 수 있다. 사용자는 관심이 있는 인물의 위치를 마우스를 사용하여 예를 들어 눈의 위치를 클릭함으로써 표시할 수 있다. 인물 검출기(110)가 인물을 검출할 때, 인물의 위치는 사용자에게 예를 들어, 디스플레이(332) 위의 얼굴에 동그라미를 두름으로써 강조될 수 있다. 그리고 나서, 사용자는 강조된 인물에 대한 이름 또는 식별자를 제공할 수 있으며, 그것에 의하여 인물의 위치를 사용자가 제공한 라벨에 관련짓는다. 한 명을 초과하는 인물이 영상 안에 검출되었을 때, 인물의 위치는 차례로 강조될 수 있으며 사용자에게 의해 임의의 인물에 대하여 라벨이 제공될 수 있다.
- <56> (4) 관심이 있는 인물을 포함한다고 생각되는 영상 컬렉션으로부터 영상 또는 비디오에 대한 검색을 하기 위한 표시(an indication).
- <57> (5) 영상 안에 있지 않은 관심이 있는 인물의 이름 또는 식별자.
- <58> 디지털 영상 컬렉션(102)은 한 명을 초과하는 인물이 있는 하나 이상의 영상을 포함한다. 라벨은 라벨러(104)를 통해 사용자에게 제공되며, 영상이 관심이 있는 인물을 포함한다는 것을 표시한다. 관심이 있는 인물에 관련되는 특징은 특징 추출기(106)에 의해 판정되며, 이 특징은 관심이 있는 인물을 포함한다고 생각되는 컬렉션의 다른 영상을 식별하기 위해 인물 탐색기(108)에 의해 사용된다.
- <59> "태그(tag)", "캡션(caption)" 및 "주석(annotation)"이라는 용어는 "라벨(label)"이라는 용어와 동의어로 사용된다는 것을 유의하라.
- <60> 도 3은 디지털 카메라를 사용하여 관심이 있는 인물을 포함한다고 생각되는 영상을 식별하는 방법을 도시하는 순서도이다. 당업자는 본 발명을 사용하기 위한 처리 플랫폼(processing platform)이 카메라, 개인용 컴퓨터, 인터넷, 프린터 또는 이와 같은 것과 같은 네트워크를 통해 액세스되는 원격 컴퓨터 일 수 있다는 것을 인식할 것이다. 이 실시예에서, 사용자는 관심이 있는 인물을 포함하는 약간의 영상 또는 비디오를 선택하고, 시스템은 관심이 있는 인물을 포함한다고 생각되는 디지털 영상 컬렉션의 서브세트로부터 영상 또는 비디오를 판정하고 디스플레이한다. 디스플레이된 영상은 사용자에게 의해 재검토될 수 있으며, 사용자는 디스플레이된 영상이 관심이 있는 인물을 정말로 포함하는지 여부를 표시할 수 있다. 추가로, 사용자는 관심이 있는 인물의 이름을 확인하거나 제공할 수 있다. 마지막으로, 사용자로부터의 입력에 기초하여, 시스템은 관심이 있는 인물을 포함한다고 생각되는 영상의 세트를 다시 판정할 수 있다.
- <61> 블록(202)에서, 영상은 디스플레이(332) 위에 디스플레이된다. 블록(204)에서 사용자는 영상을 선택하며, 각 영상은 관심이 있는 인물을 포함한다. 선택된 영상 중 하나 이상은 관심이 있는 인물 외에도 인물을 포함한다. 예를 들어, 도 4는 각각 관심이 있는 인물을 포함하고 그 중 하나의 영상은 두 인물을 포함하는 세 개의 선택된 영상의 세트를 도시한다. 블록(206)에서, 사용자는 라벨러(104)를 통해 라벨을 제공하며, 이는 선택된 영상이 관심이 있는 인물을 포함한다고 표시하고, 영상 컬렉션으로부터의 영상 및 비디오는 관심이 있는 인물을 포함한다고 생각되는 것을 식별하기 위하여 인물 탐색기(108)에 의해 검색될 것이라는 것을 표시한다. 블록(208)에서, 인물 식별자는 데이터베이스(114)에 저장된 관련된 라벨 및 특징에 액세스하고 관심이 있는 인물을 포함한다고 생각되는 영상 및 비디오의 디지털 영상 컬렉션 서브세트(112)를 판정한다. 블록(210)에서, 디지털 영상 컬렉션 서브세트(112)는 디스플레이(332) 위에 디스플레이된다. 예를 들어, 도 5는 디지털 영상 컬렉션 서브세트(112)의 영상을 도시한다. 디지털 영상 컬렉션 서브세트는 라벨 붙여진 영상(220), 관심이 있는 인물을 포함한다고 정확하게 생각되는 영상(222) 및 관심이 있는 인물을 포함한다고 부정확하게 생각되는 영상(224)을 포함한다. 이는 현재의 얼굴 검출 및 인식 기술의 불완전한 성질의 결과이다. 블록(212)에서, 사용자는 디지털 영상 컬렉

션 서브세트(112)를 재검토하며 디지털 영상 컬렉션 서브세트(112)의 각 영상의 정확성을 표시할 수 있다. 이 정확성의 사용자 표시는 블록(214)에서 라벨러(104)를 통해 추가적인 라벨을 제공하기 위해 사용된다. 예를 들어, 사용자는 사용자 인터페이스를 통해 디지털 영상 컬렉션 서브세트(112)의 관심이 있는 사람을 포함하는 것으로 정확하게 생각되는 모든 영상 및 비디오(222)가 관심이 있는 인물을 정말로 포함한다고 표시한다. 디지털 영상 컬렉션의 각 영상 및 비디오는 그리고 나서, 만약 그것이 사용자에게 의해 제공되었다면, 관심이 있는 인물의 이름으로 라벨 붙여진다. 만약 관심이 있는 인물의 이름이 사용자에게 의해 제공되지 않았다면, 관심이 있는 인물의 이름은 라벨러(104)에 의해 몇몇 경우에서 판정될 수 있다. 디지털 영상 컬렉션 서브세트(112)의 영상 및 비디오는 관심이 있는 인물의 이름을 표시하는 라벨을 가지고 있는 것에 대하여 조사되며, 이에 대하여 인물 검출기(110)가 판정하는 것은 단지 한 명의 인물만을 포함한다. 사용자가 디지털 영상 컬렉션 서브세트(112)의 영상 및 비디오가 관심이 있는 인물을 정말로 포함한다는 것을 확인하였고 인물 검출기(110)가 단지 단일한 인물을 탐색하였기 때문에, 라벨러(104)는 라벨에 관련된 인물의 이름이 관심이 있는 인물의 이름이라고 결론지을 수 있다. 만약 인물 검출기(110)가 자동 에러-프론(automatic error-prone) 알고리즘이라면, 라벨러(104)는 만약 하나 이상의 영상 및 비디오가 인물의 이름을 포함하는 관련된 라벨을 가지고 있고 인물 검출기(110)가 단지 한 명의 인물을 탐색하였으며, 관련된 라벨의 인물의 이름이 만장일치 하지 않는다면, 투표 기법(a voting scheme)을 구현하는 것이 필요할 수 있다. 예를 들어, 디지털 영상 컬렉션 서브세트(112) 중 각각 인물 검출기(110)에 의해 검출된 한 명의 인물을 포함하는 세 개의 영상이 있고, 각 영상은 인물의 이름을 포함하는 라벨을 가지고 있으며, 이름은 "한나(Hannah)", "한나(Hannah)" 및 "할리(Holly)"라면, 라벨러(104)에 의해 투표 기법이 수행되어 인물의 이름이 "한나(Hannah)"라는 것을 판정한다. 라벨러(104)는 그리고 나서 관심이 있는 인물의 이름(예, "한나(Hannah)")을 포함하는 라벨로 디지털 영상 컬렉션 서브세트(112)의 영상 및 비디오에 라벨 붙인다. 사용자는 라벨러(104)에 의해 판정된 관심이 있는 인물의 이름을 디스플레이를 통해 재검토할 수 있다. 사용자가 디지털 영상 컬렉션 서브세트(112)의 영상 및 비디오가 관심이 있는 인물을 포함하는 것을 표시한 후, "한나로 라벨 붙이겠습니까?(Label as Hannah?)"라는 메시지가 나타나며, 사용자는 "네(yes)"를 누름으로써 관심이 있는 인물의 판정된 이름을 승인하거나 "아니오(no)"를 누름으로써 관심이 있는 인물에 대하여 다른 이름을 입력할 수 있다. 만약 라벨러(104)가 관심이 있는 인물의 이름을 판정할 수 없다면, 현재 사용되지 않은 식별자가 관심이 있는 인물에 할당되며(예, "인물 12(Person 12)"), 디지털 영상 컬렉션 서브세트(112)의 영상 및 비디오는 라벨러(104)에 의해서 그에 따라서 라벨 붙여진다.

- <62> 이와 달리, 라벨러(104)는 관심이 있는 인물에 대하여 몇몇 후보 라벨을 판정할 수 있다. 후보 라벨은 목록의 형태로 사용자에게 디스플레이될 수 있다. 후보 라벨의 목록은 과거에 사용된 라벨의 목록 또는 관심이 있는 현재의 특정 인물에 대한 가장 유망한 라벨의 목록일 수 있다. 사용자는 그리고 나서 목록으로부터 관심이 있는 인물에 대한 원하는 라벨을 선택할 수 있다.
- <63> 이와 달리, 만약 라벨러(104)가 관심이 있는 인물의 이름을 판정할 수 없다면, "이 사람은 누구입니까?(Who is this?)"라는 메시지를 디스플레이(332) 위에 디스플레이하고 사용자가 관심이 있는 인물의 이름을 입력하는 것을 허용함으로써 사용자는 관심이 있는 인물의 이름을 입력하도록 요청받을 수 있으며, 이는 라벨러(104)에 의해 디지털 영상 컬렉션 서브세트(112)의 영상 및 비디오에 라벨 붙이기 위하여 사용될 수 있다.
- <64> 사용자는 또한, 사용자 인터페이스를 통해, 디지털 영상 컬렉션 서브세트(112)의 영상 및 비디오의 이러한 영상이 관심이 있는 인물을 포함하지 않는다는 것을 표시할 수 있다. 표시된 영상은 그리고 나서 디지털 영상 컬렉션 서브세트(112)로부터 제거되며, 남아있는 영상은 전술한 것과 같이 라벨 붙여질 수 있다. 표시된 영상은 동일한 관심이 있는 인물에 대한 앞으로의 검색에서, 관심이 있는 인물을 포함하지 않는다고 명시적으로 라벨 붙여진 영상은 사용자에게 도시되지 않도록, 이들이 관심이 있는 인물을 포함하지 않는다는 것을 표시하기 위해 라벨 붙여질 수 있다. 예를 들어, 도 6은 관심이 있는 인물을 포함한다고 부정확하게 생각된 영상이 제거된 후의 디지털 영상 컬렉션 서브세트(112)를 도시한다.
- <65> 도 7은 관심이 있는 인물을 포함한다고 생각되는 영상을 식별하기 위한 다른 방법을 도시하는 순서도이다. 이 실시예에서, 사용자는 하나 이상의 영상 또는 비디오에 라벨 붙이며, 관심이 있는 인물에 대한 검색을 개시하며, 시스템은 관심이 있는 인물을 포함한다고 생각되는 디지털 영상 컬렉션(102)의 서브세트로부터 영상 또는 비디오를 판정하고 디스플레이한다. 디스플레이된 영상은 사용자에게 의해 재검토될 수 있으며, 사용자는 디스플레이된 영상이 관심이 있는 인물을 정말로 포함하는지 여부를 표시할 수 있다. 추가로, 사용자는 관심이 있는 인물의 이름을 확인하거나 제공할 수 있다. 마지막으로, 사용자로부터의 입력에 기초하여, 시스템은 관심이 있는 인물을 포함한다고 생각되는 영상의 세트를 다시 판정할 수 있다.
- <66> 블록(202)에서, 영상은 디스플레이(322) 위에 디스플레이된다. 블록(204)에서, 사용자는 영상을 선택하고, 각

영상은 관심이 있는 인물을 포함한다. 하나 이상의 선택된 영상은 한 명을 초과하는 인물을 포함한다. 블록(206)에서, 사용자는 선택된 영상 안의 인물을 식별하기 위해 라벨러(104)를 통해 라벨을 제공한다. 바람직하게는, 라벨은 영상 또는 비디오 안의 인물의 위치를 표시하지 않는다. 바람직하게는, 라벨은 선택된 영상 또는 비디오 안의 인물 또는 인물의 이름을 표시한다. 도 8은 두 개의 선택된 영상 및 선택된 영상 각각의 안에 있는 인물의 이름을 표시하는 관련된 라벨(226)을 도시한다. 블록(207)에서, 사용자는 관심이 있는 인물에 대한 검색을 개시한다. 관심이 있는 인물은 선택된 영상 안의 인물을 라벨 붙일 때 라벨로 사용된 인물의 이름이다. 예를 들어, 사용자는 "조나(Jonah)"의 영상에 대한 검색을 개시한다. 블록(208)에서, 인물 식별자는 특징 추출기(106)로부터의 특징 및 데이터베이스(114)에 저장된 관련된 라벨에 액세스하여 관심이 있는 인물을 포함한다고 생각되는 영상 및 비디오의 디지털 영상 서브세트(112)를 판정한다. 블록(210)에서, 디지털 영상 컬렉션 서브세트(112)는 디스플레이(332) 위에 디스플레이된다. 도 9는 디지털 영상 컬렉션 서브세트(112)가 라벨 붙여진 영상(220), 관심이 있는 인물을 포함한다고 정확하게 생각되는 영상(222) 및 관심이 있는 인물을 포함한다고 부정확하게 생각되는 영상(224)을 포함한다는 것을 도시한다. 이는 현재의 얼굴 검출 및 인식 기술의 불완전한 성질의 결과이다. 블록(212)에서, 사용자는 디지털 영상 컬렉션 서브세트(112)를 재검토하며 디지털 영상 컬렉션 서브세트(112)의 각 영상의 정확성을 표시할 수 있다. 이 정확성의 사용자 표시는 블록(204)에서 라벨러(104)를 통해 추가적인 라벨을 제공하기 위해 사용된다. 예를 들어, 사용자는 사용자 인터페이스를 통해 디지털 영상 컬렉션 서브세트(112)의 관심이 있는 사람을 포함하는 것으로 정확하게 생각되는 모든 영상 및 비디오(222)가 관심이 있는 인물을 정말로 포함한다고 표시한다. 사용자는 또한 사용자 인터페이스를 통해 디지털 영상 컬렉션 서브세트(112)의 영상 및 비디오의 이러한 영상이 관심이 있는 인물을 포함하지 않는다는 것을 표시할 수 있다. 표시된 영상은 그리고 나서 디지털 영상 컬렉션 서브세트(112)로부터 제거되며, 남아있는 영상은 전술한 것과 같이 라벨 붙여질 수 있다. 디지털 영상 컬렉션 서브세트(112)의 각 영상 및 비디오는 그리고 나서 관심이 있는 인물의 이름으로 라벨 붙여진다. 사용자는 디스플레이를 통해 라벨러(104)에 의해 판정된 관심이 있는 인물의 이름을 재검토할 수 있다. 사용자가 디지털 영상 컬렉션 서브세트(112)의 영상 및 비디오가 관심이 있는 인물을 포함한다고 표시한 후, "조나로 라벨 붙이겠습니까?(Label as Jonah?)"라는 메시지가 나타나며, 사용자는 "네(yes)"를 누름으로써 관심이 있는 인물의 판정된 이름을 승인하거나 "아니오(no)"를 누름으로써 관심이 있는 인물에 대하여 다른 이름을 입력할 수 있다. 도 10은 사용자가 관심이 있는 인물을 포함한다고 부정확하게 생각되는 영상을 제거한 후의 디지털 영상 컬렉션 서브세트(112) 및 사용자에게 의해 재검토된 영상에 라벨 붙이기 위해 사용된 자동으로 생성된 라벨(228)을 도시한다.

- <67> 관심이 있는 인물 및 영상 또는 비디오가 해당 기술에서 알려진 임의의 사용자 인터페이스에 의해 선택될 수 있다는 것을 유의하라. 예를 들어, 만약 디스플레이(332)가 터치 감지 디스플레이(a touch sensitive display)라면, 관심이 있는 인물의 대략적인 위치는 사용자가 디스플레이(332)를 터치하는 위치를 판정함으로써 탐색될 수 있다.
- <68> 도 11은 도 2로부터의 특징 추출기(106)를 매우 자세히 설명한다. 특징 추출기(106)는 디지털 영상 컬렉션의 영상 및 비디오로부터 인물에 관련된 특징을 판정한다. 이 특징은 그리고 나서 관심이 있는 인물을 포함한다고 생각되는 디지털 영상 컬렉션의 영상 또는 비디오를 탐색하기 위해 인물 탐색기(108)에 의해 사용된다. 특징 추출기(106)는 인물에 관련된 두 유형의 특징을 판정한다. 글로벌 특징 검출기(global feature detector)(242)는 글로벌 특징(global featurer)(246)을 판정한다. 글로벌 특징(246)은 영상 또는 비디오 안의 개인(individual)의 신원(identity) 또는 위치(position)에 독립적인 특징이다. 예를들어, 사진사의 신원은 영상 또는 비디오 안에 얼마나 많은 인물이 있는지에 관계없이 일정하고 마찬가지로 인물의 위치 및 신원에 독립적이기 때문에 사진사의 신원은 글로벌 특징이다.
- <69> 추가적인 글로벌 특징(246)은 다음을 포함한다.
- <70> 영상/비디오 파일 이름.
- <71> 영상 비디오 포착 시간. 영상 포착 시간은 시간상으로 정확한 분(precise minute in time)일 수 있다(예, 2004년 3월 27일 오전 10:17). 또는 영상 포착 시간은 보다 덜 정확할 수 있다(예, 2004년 또는 2004년 3월). 영상 포착 시간은 확률 분포 함수(a probability distribution function)일 수 있다(예, 95% 신뢰도를 갖는 2004년 3월 27일 +/-2일). 종종 포착 시간은 디지털 영상 또는 비디오의 파일 헤더(file header) 안에 임베드 된다. 예를 들어, EXIF 영상 포맷(www.exif.org에서 설명되는)은 영상 또는 비디오 포착 디바이스가 영상 또는 비디오와 관련된 정보를 파일 헤더 안에 저장하는 것을 허용한다. "날짜/시간(Date/Time)" 엔트리(entry)는 영상이 포착된 날짜 및 시간에 관련된다. 몇몇 경우에서, 필름 스캔으로부터 유래된 디지털 영상 또는 비디오 및 영상 포착 시간은 통상적으로 영상의 낮은 쪽 왼쪽 구석 안에 있는 영상(포착 시간에서 종종 행해지는 것과 같이) 영역 안

에 프린트된 날짜를 검출함으로써 판정된다. 사진이 프린트되는 날짜는 종종 프린트의 뒷면에 프린트된다. 이와 달리, 몇몇 필름 시스템은 포착 날짜와 같은 정보를 저장하기 위하여 필름 안에 자기 층(a magnetic layer)을 포함한다.

- <72> 포착 조건 메타데이터(예, 플래시 파이어 정보(flash fire information), 셔터 속도(shutter speed), 구경(aperture), ISO, 장면 밝기(scene brightness) 등).
- <73> 지리적 위치. 위치는 바람직하게는 위도 및 경도 단위로 저장된다.
- <74> 장면 환경 정보. 장면 환경 정보는 인물을 포함하고 있지 않은 영역 안의 영상 또는 비디오의 픽셀 값으로부터 파생된 정보이다. 예를 들어, 영상 또는 비디오 안의 인물이 없는 영역(non-people region)의 평균 값(mean value)이 장면 환경 정보의 예이다. 장면 환경 정보의 다른 예는 텍스처 샘플(texture samples)이다(예, 영상 안의 벽지 영역으로부터의 픽셀 값의 샘플링).
- <75> 지리적 위치 및 장면 환경 정보는 관련된 영상 안에서 인물의 신원에 대한 중요한 단서이다. 예를 들어, 사진사의 할머니 집의 방문은 할머니가 사진 찍힌 유일한 위치가 될 수 있다. 유사한 지리적 위치 및 환경이 있는 두 개의 영상이 포착되었을 때, 두 영상 안에서 검출된 인물도 마찬가지로 동일할 것이다.
- <76> 장면 환경 정보는 두 영상을 등록하기 위해 인물 검출기(110)에 의해 사용될 수 있다. 이는 사진 찍히는 인물이 가장 움직이지 않는 상태이나, 카메라가 연속되는 사진 사이에서 조금 움직일 때 유용하다. 장면 환경 정보는 두 영상을 등록하기 위해 사용되며, 그것에 의하여 두 프레임 안의 인물의 위치를 정렬한다. 이 정렬은 두 인물이 시간상으로 가까운 두 포착된 영상 안에서 동일한 위치를 가지고 있으며 등록 되었을 때, 두 인물이 동일 개인(individual)일 가능성이 높기 때문에, 인물 탐색기(108)에 의해 사용된다.
- <77> 로컬 특징 검출기(local feature detector)(240)는 로컬 특징(local feature)(244)을 계산한다. 로컬 특징은 영상 또는 비디오 안의 인물의 겉모습(appearance)에 직접 관련되는 특징이다. 영상 또는 비디오 안의 인물에 대한 이러한 특징의 계산은 인물의 위치의 지식을 요구한다. 로컬 특징 검출기(240)는 인물 검출기(110) 또는 데이터베이스(114) 둘 중 하나 또는 모두로부터 영상 또는 비디오 안에 있는 인물의 위치에 관련되는 정보를 전달받는다. 인물 검출기(110)는 인물의 윤곽을 그리고, 눈 위치를 표시하는 것 또는 이와 같은 것을 통해 사용자가 영상 및 비디오 안의 인물의 위치를 입력하는 수동 동작(a manual operation)일 수 있다. 바람직하게, 인물 검출기(110)는 얼굴 검출 알고리즘을 구현한다. 인간 얼굴 검출을 위한 방법은 디지털 영상 처리의 해당 기술에 잘 알려져 있다. 예를 들어, 영상 안의 인간 얼굴 탐색을 위한 얼굴 검출 방법이 다음 논문, 존스, M.J.(Jones, M.J.), 비올라, P.(Viola, P.)의 "패스트 멀티-뷰 얼굴 검출(Fast Multi-view Face Detection)", *IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, June 2003 에 설명되어 있다.
- <78> 효과적인, 디지털 영상 및 비디오에 관련된 영상 포착 시간에 기초한 인물 검출기(110)가, 도 12a를 참조하여 설명되어 있다. 디지털 영상 컬렉션(102)의 영상 및 비디오는 존스(Jones) 및 비올라(Viola)에 의한 전술한 얼굴 검출기와 같은 얼굴 검출기(270)에 의해 분석된다. 얼굴 검출기는 잘못된 검출을 최소화하면서 검출된 인물(274)을 제공하기 위하여 조정된다. 그 결과, 영상 안의 많은 인물은 검출되지 않는다. 이는 예를 들어, 카메라에 그들의 등이 있거나, 또는 얼굴을 손으로 가린 것의 결과일 수 있다. 얼굴 검출기(270)로부터 검출된 얼굴 및 디지털 영상 컬렉션(102)은 얼굴 인식기(270)에 의해 놓쳐진 인물을 포함하는 영상을 찾기 위해 포착 시간 분석기(capture time analyzer)(272)로 전달된다. 포착 시간 분석기(272)는 두 개의 영상이 시간상 매우 가까이 포착되었을 때, 만약 한 개인이 하나의 영상에 나타났다면, 그 또는 그녀는 다른 영상에도 마찬가지로 나타날 것 같다는 생각 위에서 동작한다. 사실, 이 관계는 영상 안의 인물의 신원이 알려져 있을 때 영상의 큰 컬렉션을 분석함으로써 매우 좋은 정확성으로 판정될 수 있다. 비디오를 처리하기 위해, 얼굴 추적 기술이 비디오의 프레임에 걸쳐서 인물의 위치를 탐색하기 위해 사용된다. 비디오 안의 얼굴 추적의 하나의 방법은 미국 특허 제 6,770,999호에 설명되어 있으며, 여기에서는 운동 분석이 비디오 안의 얼굴을 추적하기 위해 사용된다.
- <79> 도 12b는 포착 시간 분석기(272)에 의해 사용되는 관계의 도면(plot)을 도시한다. 도면은 제 1 영상에 인물이 나타났을 때, 제 2 영상에 인물이 나타날 확률을, 영상 사이의 영상 포착 시간의 차이의 함수로 도시한다. 예상된 것과 같이, 두 영상이 빠르게 연속적으로 포착되었을 때, 하나의 영상에는 인물이 나타나고 다른 영상에는 나타나지 않을 가능성은 매우 낮다.
- <80> 포착 시간 분석기(272)는 디지털 영상 컬렉션(110)의 영상 및 비디오를 조사한다. 주어진 영상 안에서 얼굴 검출기(270)에 의해 얼굴이 검출될 때, 동일한 인물이 다른 영상에 나타날 확률은 도 12b에 도시된 관계를 사용하여 계산된다.

<81> 예를 들어, 얼굴 검출기(270)가 하나의 영상 안에서 두 개의 얼굴을 검출하였고, 단지 1초 후에 포착된 제 2 영상에서는 얼굴 검출기(270)가 단지 하나의 얼굴만 탐색하였다고 가정하자. 제 1 영상으로부터 검출된 얼굴이 확실하게 명확하다고 가정하면, 제 2 영상 또한 두 개의 얼굴을 포함할 확률이 매우 높으나(0.99*0.99), 얼굴 검출기(270)에 의해 단지 하나의 얼굴이 탐색되었다. 제 2 영상에 대하여 검출된 인물(274)은 얼굴 검출기(270)에 의해 탐색된 하나의 얼굴이며, 0.98의 신뢰도를 가지는 제 2 얼굴이다. 제 2 얼굴의 위치는 알려지지 않으나, 포착 시간 차이가 작을 때, 카메라 또는 사진 찍히는 인물 둘 중 어느 하나도 빨리 움직이는 경향이 있지 않기 때문에, 추정될 수 있다. 따라서, 제 2 영상 안에서의 제 2 얼굴의 위치가 포착 시간 분석기(272)에 의해 추정된다. 예를 들어, 개인이 두 개의 영상에 나타날 때, 상대적인 얼굴 크기(더 큰 얼굴에 대한 더 작은 얼굴의 크기의 비)가 조사될 수 있다. 동일한 인물을 포함하는 두 영상의 포착 시간이 매우 작을 때, 사진사, 사진 찍히는 사람 및 카메라 설정이 거의 일정하기 때문에, 상대적인 얼굴 크기는 보통 1 근처로 된다. 상대적인 얼굴 크기의 하한(a lower limit)이 도 12c에 영상 포착 시간의 차이의 함수로 도시된다. 이 스케일링 팩터(scaling factor)는 제 1 영상 안의 얼굴의 알려진 얼굴 위치와 함께 제 2 영상 안에 얼굴이 나타나는 영역을 추정하기 위하여 사용될 수 있다.

<82> 포착 시간 분석기(272)에 의해 사용되는 방법은 또한 인물 탐색기(108)에 의해 특정 영상 또는 비디오 안에 관심이 있는 인물이 있을 가능성을 판정하기 위해 사용될 수 있다는 것을 유의하라.

<83> 또한, 데이터베이스(114)는 도 2의 라벨러(104)로부터 라벨과 관련된 정보를 저장한다. 라벨이 인물과 관련된 위치 정보를 포함할 때, 로컬 특징 검출기(240)는 인물과 관련된 로컬 특징(244)을 판정할 수 있다.

<84> 일단 인물의 위치가 알려지면, 로컬 특징 검출기(240)는 인물과 관련된 로컬 특징(244)을 검출할 수 있다. 일단 얼굴 위치가 알려지면, 얼굴 특징(예, 눈, 코, 입 등)이 또한 유일레(Yuille) 등의 "변형할 수 있는 템플릿을 사용한 얼굴로부터의 특징 추출(Feature Extraction from Faces Using Deformable Templates)", *Int. Journal of Comp. Vis.*, Vol. 8, Iss. 2, 1992, pp. 99-111에 설명된 것과 같은 잘 알려진 방법을 사용하여 국한될 수 있다. 저자는 입, 눈 및 홍채(iris)/공막(sclera) 경계를 알아내기 위한 템플릿 정합(template matching)과 함께 에너지 최소화를 사용하는 방법을 설명한다. 얼굴 특징은 또한 T. F. 쿠테스(T. F. Cootes) 및 C. J. 테일러(C. J. Taylor)의 "강제적인 동적 겹모습 모델(Constrained active appearance models)", *8th International Conference on Computer Vision*, 1 권, 748-754 페이지, IEEE Computer Society Press, July 2001에 의해 설명된 것과 같이 동적 겹모습 모델(active appearance models)을 사용하여 탐색될 수 있다. 바람직한 실시예에서, 보린(Bolin) 및 첸(Chen)에 의한 Proceedings of IS&T PICS conference, 2002에서의 "인물 사진 영상에 대한 자동 얼굴 특징 탐색 시스템(An automatic facial feature finding system for portrait images)"에 설명된, 인간 얼굴의 동적 형상 모델(an active shape model)에 기초하여 얼굴 특징 점(facial feature points)을 알아내는 방법이 사용된다.

<85> 로컬 특징(244)은 인물의 양적인 설명이다. 바람직하게는, 인물 탐색기와 특징 추출기(106)는 각 검출된 인물에 대하여 하나의 세트의 로컬 특징(244) 및 하나의 세트의 글로벌 특징(246)을 출력한다. 바람직하게는 로컬 특징(244)은 특정 얼굴 특징과 관련된 82개의 특징 점의 위치에 기초하여, 전술한 쿠테스(Cootes) 등의 동적 겹모습 모델과 유사한 방법을 사용하여 탐색된다. 얼굴의 영상에 대한 로컬 특징 점의 시각적 표시가 도 12d에 예로 도시되어 있다. 로컬 특징은 또한 특정 특징 점 사이의 거리 또는 특정 특징 점의 세트를 연결하는 선에 의해 형성되는 각도 또는 얼굴 겹모습의 변화성(variability)을 설명하는 주요한 구성 요소 위로의 특징 점의 투영의 계수일 수 있다.

<86> 사용된 특징은 표 1에 열거되어 있으며, 그들의 계산은 도 12d에서 번호가 매겨진 도시된 얼굴 위의 점을 참조

$$\sum_{j=n}^{m-1} \|P_n - P(n+1)\|$$

한다. $Arc(P_n, P_m)$ 은 $\|P_n - P_m\|$ 에 의해 규정되며, $\|P_n - P_m\|$ 은 특징 점 n 과 m 사이의 유클리드 거리(Euclidean distance)를 가리킨다. 호-길이(arc-length) 특징은 상이한 얼굴 크기에 걸쳐서 표준화하기 위하여 내안 거리(inter-ocular distance)에 의해 나누어진다. 점 PC는 점 0 및 1의 중심에 위치한 점(즉, 눈 사이에 정확히 위치한 점)이다. 여기에 사용된 얼굴 측정치는 성별, 나이, 매력 및 민족성을 판단하는데 적절한 것으로 보이는 인간 얼굴의 인체 측정학적 측정으로부터 파생된다(팔카스(Farkas)(편집(Ed.))의 "머리 및 얼굴의 인체 측정학(Anthropometry of the Head and Face)", 2 판, Laven Press, New York, 1994 참조).

<87> 표 1 : 할당량 특징 목록(List of Ration Features)

이름	분자	분모
눈에서 코까지(Eye-to-nose)/눈에서 입까지(Eye-to-mouth)	PC-P2	PC-P32
눈에서 입까지(Eye-to-mouth)/눈에서 턱끝까지(Eye-to-chin)	PC-P32	PC-P75
머리에서 턱끝까지(Head-to-chin)/눈에서 입까지(Eye-to-mouth)	P62-P75	PC-P32
머리에서 눈까지(Head-to-eye)/눈에서 턱끝까지(Eye-to-chin)	P62-PC	PC-P75
머리에서 눈까지(Head-to-eye)/눈에서 입까지(Eye-to-mouth)	P62-PC	PC-P32
코에서 턱끝까지(Nose-to-chin)/눈에서 턱끝까지(Eye-to-chin)	P38-P75	PC-P75
입에서 턱끝까지(Mouth-to-chin)/눈에서 턱끝까지(Eye-to-chin)	P35-P75	PC-P75
머리에서 코까지(Head-to-chin)/코에서 턱끝까지(Nose-to-chin)	P62-P2	P2-P75
입에서 턱끝까지(Mouth-to-chin)/코에서 턱끝까지(Nose-to-chin)	P35-P75	P2-P75
턱 너비(Jaw width)/얼굴 너비(Face width)	P78-P72	P56-P68
눈-간격(Eye-spacing)/코 너비(Nose width)	P07-P13	P37-P39
입에서 턱끝까지(Mouse-to-chin)/턱 너비(Jaw width)	P35-P75	P78-P72

<89> 표 2 : 호 길이 특징 목록(List of Arc Length Features)

이름	계산
큰 턱의 호(Mandibular arc)	Arc(P69, P81)
완와위의 호(Supra-orbital arc)	(P56-P40) + Int(P40, P44) + (P44-P48) + Arc(P48, P52) + (P52-P68)
윗-입술 호(Upper-lip arc)	Arc(P23, P27)
아랫-입술 호(Lower-lip arc)	Arc(P27, P30) + (P30 - P23)

- <91> 컬러 신호(color cues)는 일단 인물 및 얼굴 특징이 인물 탐색기(108)에 의해 알아내어지면, 디지털 영상 또는 비디오로부터 쉽게 추출된다.
- <92> 이와 달리, 상이한 로컬 특징 또한 사용될 수 있다. 예를 들어, 일 실시예는 M. 투르크(M. Turk) 및 A. 펜트랜드(A. Pentland)에 의해 "인식을 위한 아이겐페이스(Eigenfaces for Recognition)", *Journal of Cognitive Neuroscience, Vol 3, No. 1. 71-86, 1991* 에서 설명된 얼굴 유사성 메트릭(facial similarity metric)에 기초할 수 있다. 얼굴 기술자(facial descriptor)는 얼굴의 영상을 얼굴 겉모습의 변화성을 설명하는 주요한 구성 요소 함수의 세트 위에 얼굴의 영상을 투영함으로써 얻어진다. 임의의 두 얼굴 사이의 유사성은 함수의 동일한 세트 위에 각 얼굴을 투영함으로써 얻어지는 특징의 유클리드 거리를 계산함으로써 측정된다.
- <93> 로컬 특징(244)은 아이겐페이스(Eigenfaces), 얼굴 측정(facial measurements), 컬러/텍스처 정보(color/texture information), 잔물결 특징(wavelet features) 등과 같은 몇몇 다른 특징 유형의 조합을 포함할 수 있다.
- <94> 이와 달리, 로컬 특징(244)은 눈 컬러, 피부 컬러, 얼굴 형상, 안경의 존재, 옷의 설명, 머리카락의 설명 등과 같은 정량화 할 수 있는 기술자로 추가적으로 표시될 수 있다.
- <95> 예를 들어, 위스콧(Wiskott)은 "얼굴 분석을 위한 판톰 얼굴(Phantom Faces for Face Analysis)", *Pattern Recognition, Vol. 30, No. 6, pp. 837-846, 1997*에서 얼굴 위의 안경의 존재를 검출하는 방법을 설명한다. 로컬 특징은 안경의 형상 및 존재와 관련된 정보를 포함한다.
- <96> 도 12e는 얼굴 검출기에 의해 생성된 눈 위치에 기초하여 얼굴 영역(face region)(282), 옷 영역(clothing

region)(284) 및 배경 영역(background region)(286)으로 가정된 영상 안의 영역을 도시한다. 크기는 내안 거리(inter-ocular distance) 또는 IOD(왼쪽 및 오른쪽 눈 위치 사이의 거리)에 입각하여 측정된다. 얼굴은 도시된 것과 같이 3배의 IOD에 4배의 IOD의 면적을 덮는다. 옷 영역은 5배의 IOD를 덮으며 영상의 바닥까지 연장된다. 영상 안의 남아있는 영역은 배경으로 취급된다. 몇몇 옷 영역이 다른 얼굴 및 이 얼굴에 대응하는 옷 영역에 의해 덮힐 수 있다는 것을 유의하라.

<97> 디지털 영상 컬렉션(102)의 영상 및 비디오는, 미국 특허 제6,606,411호에 따라 일정한 컬러 분포를 가지며, 따라서, 이 사진은 동일한 배경에서 찍혔을 것 같은, 서브-이벤트(sub-event) 및 이벤트(event) 안으로 밀집된다. 각 서브-이벤트에 대하여, 단일한 컬러 및 텍스처 표현이, 하나로 합쳐서 생각되는 모든 배경에 대하여 계산된다. 컬러 및 텍스처 표현 및 유사성(similarity)은 미국 특허 제 6,480,840호에서 주(Zhu) 및 메호트라(Mehrotra)에 의해 파생된다. 그들의 방법에 따라, 영상의 컬러 특징-기초 표현(color feature-based representation)은 영상의 상당한 크기의 일관되게 색칠된 영역은 지각적으로 중요하다는 가정에 기초한다. 따라서, 상당한 크기의 일관되게 색칠된 영역의 컬러는 지각적으로 중요한 컬러로 생각된다. 따라서, 모든 입력 영상에 대하여, 일관된 컬러 히스토그램이 먼저 계산되며, 일관된 컬러 히스토그램은 일관되게 색칠된 영역에 속하는 특정 컬러의 픽셀의 수의 함수이다. 만약 픽셀의 컬러가 주변 픽셀의 미리-특정된(pre-specified) 최소 수의 컬러와 같거나 유사하다면, 픽셀은 일관되게 색칠된 영역에 속하는 것으로 생각된다. 더욱이, 영상의 텍스처 특징-기초 표현(texture feature-based representation)은 각 지각적으로 상당한 텍스처는 동일한 컬러 변화(color transition(s))의 많은 수의 반복으로 구성된다는 가정에 기초한다. 따라서, 빈번히 발생하는 컬러 변화를 식별하고 그들의 텍스처 특성(texture property)을 분석함으로써, 지각적으로 상당한 텍스처가 추출되고 표시될 수 있다.

<98> 얼굴 검출기에 의해 생성된 눈 위치는 얼굴 특징 탐색(facial feature finding)을 위하여 시작 얼굴 위치(starting face position)를 초기화하는데 사용된다. 도 12f는 얼굴 위의 특징 점의 위치와 이름 붙여진 2차 특징(secondary features)이 위치할 수 있는 대응하는 영상 패치(patches)를 도시한다.

<99> 표 3은 도 12에 도시된 이 영상 패치, 머리카락 영역(hair region)(502), 앞머리카락 영역(bang region)(504), 안경 영역(eyeglass region)(506), 뺨 영역(cheek region)(508), 긴 머리카락 영역(long hair region)(510), 턱수염 영역(beard region)(512) 및 콧수염 영역(mustache region)(514)에 대한 바운딩 박스(bounding boxes)를 열거하며, P_n 은 도 12f 또는 도 12d로부터의 얼굴 점 수 n (facial point number n)을 가리키며, $[x]$ 및 $[y]$ 는 점의 x 및 y 좌표를 가리킨다. $(P_n - P_m)$ 은 점 n 및 m 사이의 유클리드 거리(Euclidean distance)이다. "뺨(cheek)" 및 "머리카락(hair)" 패치(patches)는 얼굴의 특징 없는 영역(feature-less region) 및 인물의 머리카락(hair)을 각각 묘사하는 기준 패치(reference patches)로 취급된다(표에서 [R]로 표시됨). 2차 특징(secondary features)은 2차 특징을 포함하는 포텐셜(potential) 패치와 적절한 기준 패치 사이의 그레이-스케일(gray-scale) 히스토그램 차이로 계산된다. 왼쪽 및 오른쪽 패치는 각 2차 특징에 대한 히스토그램을 생성하기 위하여 조합된다. 히스토그램은 비교되는 패치의 상대적인 크기가 계산된 차이의 요소가 아니도록 하기 위하여 픽셀의 수에 의해서 정규화된다. 2차 특징은 마이너리 특징으로 취급된다 - 이들은 존재하거나(present) 결여된다(absent). 문턱값(a threshold)은 2차 특징이 존재하는지 여부를 확인하기 위하여 사용된다. 표 4는 검출될 각 2차 특징에 대하여 사용되는 히스토그램 차이를 도시하는 표를 제공한다.

<100> 표 3. 얼굴 특징 영역의 바운딩 박스

	바운딩 박스(Bounding Box)			
	x-시작(x-start)	y-시작(y-start)	폭(width)	높이(height)
뺨[R] (오른쪽)	$P80[x] + 1/3$ ($P37 - P80$)	$Mean(P80[y], P81[y])$	$2/3(P37-P80)$	$P79-P80$
뺨[R] (왼쪽)	$P39[x]$	$Mean(P69[y], P70[y])$	$2/3(P39-P70)$	$P70-P69$
머리카락[R]	$P61[x]$	$P62[y]$ -높이(height)	$P63-P61$	$P68-P17$
긴 머리카락 (왼쪽)	$P56[x]$ - $2*폭(width)$	$P56[y]$	$P56-P3$	$P56-P79$
긴 머리카락 (오른쪽)	$P68[x]$ +폭(width)	$P68[y]$	$P68-P17$	$P71-P68$
안경 (왼쪽)	$P56[x]+1/3$ ($P7-P56$)	$Mean(P56[y], P81[y])$	$2/3(P7-P56)$	$1/2 (P56-P81)$

안경 (오른쪽)	P13[x]	Mean(P68[y],P69[y])	2/3(P13-P68)	1/2 (P69-P68)
앞머리카락	P60[x]	Mean(P60[y],P64[y])	P64-P60	2/3(P42-P60)
콧수염	P23[x]	P38[y]	P27-P23	P38-P25
턱수염	Mean(P30[x], P76[x])	Mean(P75[y],P35[y])	Mean(P28- P30,P74-P76)	1/2 (P75-P35)

표 4. 2차 특징에 대한 히스토그램 차이

특징	히스토그램 차이 테스트
긴 머리카락	긴 머리카락 - 머리카락 < 문턱값
안경	안경 - 뺨 > 문턱값
앞머리카락	앞머리카락 - 뺨 > 문턱값
콧수염	콧수염 - 뺨 > 문턱값
턱수염	턱수염 - 뺨 > 문턱값

다시 도 11을 참조하면, 글로벌 특징(246) 및 로컬 특징(244)은 데이터베이스(114)에 저장된다. 영상 안의 모든 인물에 관련된 글로벌 특징은 F_G 로 표시된다. 영상 안의 N 명의 인물에 관련된 로컬 특징의 N 세트는 $F_{L0}, F_{L1}, \dots, F_{LN-1}$ 으로 표시된다. 영상 안의 인물 n 에 대한 특징의 완전한 세트는 F_n 으로 표시되며 글로벌 특징 F_G 및 로컬 특징 F_{Ln} 을 포함한다. 영상에 관련된 M 라벨은 L_0, L_1, \dots, L_{M-1} 으로 표시된다. 라벨이 인물의 위치를 포함하지 않을 때, 어떤 라벨이 영상 또는 비디오 안의 인물을 표시하는 어떤 특징 세트에 관련되는지 아는 것에 모호함(ambiguity)이 있다. 예를 들어, 영상 안의 두 인물과 두 라벨을 설명하는 두 세트의 특징이 있을 때, 어떤 특징이 어떤 라벨에 속하는지 명확하지 않다. 인물 탐색기(108)는 라벨을 로컬 특징의 세트와 정합시키는 이 강제적인 분류 문제를 해결하며, 여기에서 라벨 및 로컬 특징은 단일한 영상에 관련된다. 임의의 수의 라벨 및 로컬 특징이 있을 수 있으며, 심지어 각각이 상이할 수도 있다.

여기에 데이터베이스(114)의 영상에 관련된 특징 및 라벨의 예시적인 엔트리(example entry)가 있다.

영상 101_346.JPG

라벨 L_0 : 한나(Hannah)

라벨 L_1 : 조나(Jonah)

특징 F_0 :

글로벌 특징 F_G :

포착 시간 : 2005년 8월 7일, 6:41 PM 동부표준시(EST)

플래시 파이어 : 아니오

셔터 스피드 : 1/724 초

카메라 모델 : 코닥 C360 줌 디지털 카메라(Kodak C360 Zoom Digital Camera)

구경(Aperture) : F/2.7

환경(Environment) :

로컬 특징 F_{L0} :

<118> 위치 : 왼쪽 눈 : [1400 198] 오른쪽 눈 : [1548 202]

<119> $C_0 = [-0.8, -0.01]'$;

<120> 안경 : 없음

<121> 관련된 라벨 : 알려지지 않음

<122> 특징 F_I :

<123> 글로벌 특징 F_G :

<124> 포착 시간 : 2005년 8월 7일, 6:41 PM 동부표준시(EST)

<125> 플래시 파이어 : 아니오

<126> 셔터 스피드 : 1/724 초

<127> 카메라 모델 : 코닥 C360 줌 디지털 카메라(Kodak C360 Zoom Digital Camera)

<128> 구경(aperture) : F/2.7

<129> 환경(Environment) :

<130> 로컬 특징 F_{LI} :

<131> 위치 : 왼쪽 눈 : [810 192] 오른쪽 눈 : [956 190]

<132> $C_I = [0.06, 0.26]'$;

<133> 안경 : 없음

<134> 관련된 라벨 : 알려지지 않음

<135> 도 13은 도 2의 인물 탐색기(108)를 매우 자세히 설명한다. 인물 식별자(250)는 데이터베이스(114)의 라벨 및 특징을 고려하여, 인물의 위치를 포함하지 않는 라벨로 라벨 붙여진, 영상 안의 인물의 신원을 판정(즉, 관련된 특징의 세트를 판정)한다. 인물 식별자(250)는 특징 추출기(106)로부터의 특징을 라벨러(104)로부터의 라벨과 관련시키며, 그것에 의하여 영상 또는 비디오 안의 인물을 식별한다. 인물 식별자(250)는 데이터베이스로부터 특징을 업데이트하며 데이터베이스(114)에 저장된 수정된 특징(254)을 생성한다. 예로서, 도 8에 도시된 영상을 생각하자. 제 1 영상(260)은 두 명의 인물을 포함하며, 이들은 라벨(226)에 따르면 한나와 조나이다. 그러나 라벨이 위치를 포함하고 있지 않으므로 어떤 인물이 한나이며 어떤 인물이 조나인지는 알려져 있지 않다. 제 2 영상(262)는 하나로 라벨 붙여져 있다. 단지 한 명의 인물이 있으므로, 이 인물은 높은 신뢰도로 하나로 식별될 수 있다. 인물 식별자(250)는 제 2 영상(262)으로부터의 한나에 관련된 특징을 사용하고 제 1 영상(260) 안의 인물의 특징을 비교함으로써 제 1 영상(260) 안의 인물의 신원을 판정한다. 인물(266)은 제 2 영상(262) 안에서 하나로 식별된 인물(264)의 특징과 유사한 특징을 가진다. 인물 식별자(250)는 높은 신뢰도로 제 1 영상(260) 안의 인물(266)이 한나이며, 배제(elimination)에 의해 인물(268)이 조나라고 결론지을 수 있다. 제 1 영상(260)

에 대한 라벨(226) 한나는 영상에 대한 글로벌 특징 F_G 및 인물(266)에 관련된 로컬 특징에 관련된다. 제 1 영상(260)에 대한 라벨(226) 조나는 영상에 대한 글로벌 특징 및 인물(268)에 관련된 로컬 특징에 관련된다. 인물의 신원이 판정되었으므로, 사용자는 적절한 특징을 사용하여 한나 또는 조나 둘 중 하나에 대한 검색을 개시할 수 있다.

<136> 일반적으로 말해서, 인물 식별자(250)는 분류 문제를 해결한다. 문제는 위치 정보를 가지고 있지 않은 라벨을 로컬 특징과 관련시키는 것이며, 여기에서 라벨 및 로컬 특징은 모두 동일한 영상에 관련된다. 이 문제를 해결하기 위한 알고리즘은 인물 식별자(250)에 의해 구현된다. 도 14는 디지털 영상 컬렉션으로부터 계산된 실제의 로컬 특징의 표시를 도시한다. 로컬 특징의 15개의 세트의 위치가 도면 위에 마크되어 있다. 마크(mark)를 표시하기 위하여 사용된 기호(symbol)는 로컬 특징, 한나에 대한 "x", 조나에 대한 "+", 홀리에 대한 "*" 및 앤디에

대한 \square (박스)와 관련된 인물의 진정한 신원을 표시한다. 로컬 특징의 각 세트는 영상에 할당된 어떠한 라벨에도 관련될 수 있다. 도면 위에 마크된 로컬 특징의 거의 각각의 세트는 로컬 특징, 앤디에 대한 "A", 한나에 대한 "H", 조나에 대한 "J" 및 홀리에 대한 "O"에 관련될 수 있는 가능한 라벨이다. 아래의 표는 데이터를 도시한다. 도면 위의 마크 사이의 링크(links)는 로컬 특징의 세트가 동일한 영상으로부터의 것이라는 것을 표시한다. 로컬 특징을 라벨에 할당하기 위한 알고리즘은 데이터 점의 공동 변화(collective variance)(즉, 각 인물에게 할당된 데이터 점의 확산(spread)의 합)을 최소화하는, 라벨에의 로컬 특징의 할당을 탐색함으로써 작동한다. 라벨에의 로컬 특징의 할당은 라벨이 각 영상에 대해 단지 한 번(즉, 링크에 의해 연결된 데이터 점의 각 세트에 대하여 한 번)만 사용될 수 있다는 제한에 영향을 받는다. 바람직하게는, 공동 변화(collective variance)는 데이터 점으로부터 동일한 개인에게 할당된 모든 데이터 점의 중심까지의 제곱 거리의 각 점에 대한 합으로 계산된다.

<137> 로컬 특징을 분류하기 위한 알고리즘은 식에 의해 요약될 수 있다.

$$\min_{d_j} \sum_j (c_{d_j} - f_j)^T (c_{d_j} - f_j)$$

<138> 여기서,

<139> f_j 는 로컬 특징의 j 번째 세트를 표시하며,

<140> d_j 는 로컬 특징의 j 번째 세트가 할당된 클래스(class)(즉, 개인의 신원)를 표시하며,

<141> c_{d_j} 는 로컬 특징의 j 번째 세트가 할당된 클래스의 중심을 표시한다.

<142> 로컬 특징의 j 번째 세트 각각에 대한 클래스의 할당을 선택함으로써 표현은 최소화될 수 있다.

<143> 이 식에서, 유클리드 거리(Euclidean distance)가 측정된다. 당업자는 마할라노비스(Mahalanobis) 거리와 같은 많은 상이한 거리 측정 또는 동일한 클래스에 할당된 다른 데이터 점과 현재 데이터 점 사이의 최소 거리가 또한 사용될 수 있다는 것을 인식할 것이다.

<144> 이 알고리즘은 예의 모든 15개의 로컬 특징을 정확한 라벨에 정확히 관련시킨다. 비록 이 예에서 각 영상 안의 로컬 특징의 수와 라벨의 수가 각 영상의 경우에서 동일하였지만, 이는 인물 인식자(250)에 의해 사용된 알고리즘이 유용한 것이 되기 위하여 필수적인 것은 아니다. 예를 들어, 사용자는 세 명의 인물을 포함하고 이로부터 로컬 특징의 세 개의 세트가 과생된 영상에 대하여 단지 두 개의 라벨만을 제공할 수 있다.

<145> 몇몇 경우에서, 인물 식별자(250)로부터의 수정된 특징(254)은 데이터베이스(114)로부터 생성하기에 직접적이다. 예를 들어, 데이터베이스가 단지 글로벌 특징만을 포함하고 로컬 특징은 포함하지 않을 때, 각 라벨에 관련된 특징(라벨이 위치 정보를 포함하든 포함하지 않든)은 독립적일 것이다. 예를 들어, 만약 유일한 특징이 포착 시간이라면, 영상에 관련된 각 라벨은 영상 포착 시간에 관련된다. 또한, 만약 라벨이 위치 정보를 포함한다면, 특징을 라벨에 관련시키는 것은, 특징이 로컬 특징을 포함하지 않아서 동일한 특징이 각 라벨에 관련되기 때문에, 또는 특징이 로컬 특징을 포함하고 로컬 특징이 계산된 영상 영역의 위치가 특징을 라벨에 관련시키는데 사용되기 때문에(근접도에 기초함), 쉽다.

<146> 인물 분류기(256)는 관심이 있는 인물을 포함한다고 생각되는 영상 및 비디오의 디지털 영상 컬렉션 서브세트(112)를 판정하기 위해 수정된 특징(254) 및 관심이 있는 인물의 신원(252)을 사용한다. 수정된 특징(254)은 관련된 라벨(라벨 붙여진 특징으로 알려진)을 가지는 몇몇 특징을 포함한다. 다른 특징(라벨 붙여지지 않은 특징으로 알려진)은 관련된 라벨을 가지지 않는다(예, 라벨러(104)에 의해 라벨 붙여지지 않은 디지털 영상 컬렉션(102)의 모든 영상 및 비디오). 인물 분류기(256)는 라벨 붙여지지 않은 특징을 분류하기 위하여 라벨 붙여진 특징을 사용한다. 이 문제는, 비록 실제로는 매우 어렵지만, 패턴 인식 분야에서 연구된다. 임의의 분류기(classifier)가 라벨 붙여지지 않은 특징을 분류하기 위해 사용될 수 있다. 바람직하게는, 인물 분류기는 각 라벨 붙여지지 않은 특징(unlabeled features) 및 제안된 라벨에 관련된 신뢰도(confidence), 믿음(belief) 또는 확률(probability)에 대하여 제안된 라벨을 판정한다. 일반적으로, 분류기는 라벨 붙여지지 않은 특징의 특정 세트와 특징의 라벨 붙여진 세트 사이의 유사성을 고려함으로써 라벨 붙여지지 않은 특징에 라벨을 할당한다. 몇몇 분류기(예, 가우시안 최대 가능성(Gaussian Maximum Likelihood))에 의해, 단일한 개인 인물에 관련된 특

장의 라벨 붙여진 세트는 개인에 대한 걸모습의 모델을 형성하기 위해 모아진다. 디지털 영상 컬렉션 서브세트 (112)는 문턱값(threshold) T_0 를 초과하는 확률을 가지는 관련된 제안된 라벨을 가지는 영상 및 비디오의 컬렉션이며, T_0 는 $0 \leq T_0 \leq 1.0$ 의 범위에 있다. 바람직하게는, 디지털 영상 컬렉션 서브세트(112)는 또한 관심이 있는 인물의 신원(252)에 정합하는 라벨을 가지는 특징에 관련되는 영상 및 비디오를 포함한다. 디지털 영상 컬렉션 서브세트의 영상 및 비디오는 영상 및 비디오가 관심이 있는 인물이 서브세트의 위쪽에 나타나는 것을 포함하고, 관심이 있는 인물의 신원(252)에 정합하는 라벨을 구비하는 특징이 있는 영상 및 비디오만이 뒤따르는, 가장 높은 믿음(belief)을 가지도록 판정되기 위하여 정렬된다.

<148> 인물 분류기(256)는 인물의 유사성과, 그것에 의하여 인물이 동일할 가능성을 판정하기 위하여, 두 명 이상의 인물에 관련되는 특징의 세트 사이에서 유사성을 측정할 수 있다. 특징의 세트의 유사성을 측정하는 것은 특징의 서브세트의 유사성을 측정함으로써 이루어진다. 예를 들어, 로컬 특징이 옷을 설명할 때, 다음의 방법이 특징의 두 개의 세트를 비교하기 위하여 사용된다. 만약 영상 포착 시간의 차이가 작다면(즉, 몇 시간 보다 작다면) 그리고 만약 옷의 양적인 설명이 특징의 두 개의 세트 각각에서 유사하다면, 로컬 특징의 두 개의 세트가 동일한 인물에 속할 가능성은 증가한다. 만약, 추가로, 로컬 특징의 모든 세트에 대하여 옷이 매우 고유하거나 독특한 패턴을 가지고 있다면(예, 커다란 녹색, 적색 및 청색 패치의 셔츠), 관련된 인물이 동일한 개인일 가능성은 더 커진다.

<149> 옷은 상이한 방법으로 표시될 수 있다. 주(Zhu) 및 메호로트라(Mehrotra)에 의해 미국 특허 제 6,480,840호에 설명된, 유사성과 컬러 및 텍스처 표시가 하나의 가능한 방법이다. 다른 가능한 표시에서, 주(Zhu) 및 메호로트라(Mehrotra)는 특허 미국 특허 제 6,584,465호의 직물(textile)에서 탐색되는 것과 같은 패턴을 정합시키고 표시하는 목적에 사용하고자 하는 방법을 설명한다. 이 방법은 컬러 불변이며(color invariant), 특징으로서 에지 방향(edge directions)의 히스토그램을 사용한다. 이와 달리, 에지 맵(edge maps) 또는 옷 패치 영상의 푸리에 변환 계수(Fourier transform coefficients)로부터 파생된 특징은 정합을 위한 특징으로 사용될 수 있다. 에지-기초(edge-based) 또는 푸리에-기초(Fourier-based) 특징을 계산하기 전에, 에지의 주파수(frequency)가 카메라/줌으로부터의 물체의 거리에 불변하도록 만들기 위해, 패치는 동일한 크기로 정규화된다. 멀티플리кат티브(multiplicative) 요소가 계산되며, 이는 검출된 얼굴의 내안 거리를 표준 내안 거리로 변환한다. 경로 크기는 내안 거리 차이로부터 계산되므로, 옷 패치는 그리고 나서 표준-크기 얼굴(standard-size face)에 대응하도록 이 요소에 의해 서브-샘플되거나(sub-sampled) 또는 확장된다.

<150> 고유성(uniqueness) 측정은, 표 5에 도시된 것과 같이 인물에 대한 전반적인 정합 점수(match score)에의 정합(match) 또는 부정합(mismatch)의 기여를 판정하는 각 옷 패턴에 대하여 계산되며, 기여의 세기를 표시하기 위해 사용된 +또는 -의 수와 함께, +는 양의 기여를 표시하고, -는 음의 기여를 표시한다. 고유성 점수는 패턴의 고유성 및 컬러의 고유성의 합으로 계산된다. 패턴의 고유성은 패치의 푸리에 변환 안의 문턱값을 넘는 푸리에 계수의 수에 비례한다. 예를 들어, 무늬가 없는(plain) 패치 및 단일한 동등한 간격의 줄이 있는 패치는 각각 1(dc 만) 또는 2 계수를 가지며, 따라서 낮은 고유성 점수를 가진다. 패턴이 더 복잡할수록, 이를 설명하기 위해 필요한 계수의 수는 더 높아지며, 그 고유성 점수는 더 높아진다. 컬러의 고유성은 인물 영상의 커다란 데이터베이스로부터, 특정 컬러가 옷에 나타날 가능성을 학습함으로써 측정된다. 예를 들어, 흰색 셔츠를 입고 있는 인물의 가능성은 오렌지색 및 녹색 셔츠를 입고 있는 인물의 가능성보다 더 크다. 이와 달리, 신뢰할만한 가능성 통계치의 결여에서, 포화된 컬러는 더 드물며 또한 덜 모호함과 정합될 수 있기 때문에, 컬러 고유성은 그 포화도(saturation)에 기초한다. 이러한 방식으로, 영상의 포착 시간과 함께 취해진 옷 유사성 또는 비유사성은, 옷의 고유성과 마찬가지로, 인물 분류기(256)가 관심이 있는 인물을 인식하기 위한 중요한 특징이다.

<151> 옷 고유성은 인물 영상의 커다란 데이터베이스로부터 특정 옷이 나타날 가능성을 학습함으로써 측정된다. 예를 들어, 흰색 셔츠를 입고 있는 인물의 가능성은 오렌지색 및 녹색 격자무늬 셔츠를 입고 있는 인물의 가능성보다 더 크다. 이러한 방식으로, 영상의 포착 시간과 함께 취해진 옷 유사성 또는 비유사성은, 옷의 고유성과 마찬가지로, 인물 분류기(256)가 관심이 있는 인물을 인식하기 위한 중요한 특징이다.

<152> 표 5. 두 인물이 동일한 개인일 가능성에의 옷의 영향

<153>

		옷 고유성(Clothing Uniqueness)	
시간 간격 (Time interval)		보통(common)	드물(rare)

동일한 이벤트	정합	++	+++
	정합하지 않음	--	---
상이한 이벤트	정합	+	+++
	정합하지 않음	영향 없음	영향 없음

- <154> 표 5는 두 인물의 가능성이 옷의 설명을 사용함으로써 어떻게 영향을 받는지를 도시한다. 두 인물이 동일한 이벤트로부터의 영상 또는 비디오로부터일 때, 인물이 동일한 개인일 가능성은 옷이 정합하지 않을 때 큰 양으로 감소한다(---). "동일한 이벤트(same event)"는 영상이 단지 영상 포착 시간 사이에 작은 차이(즉, 몇 시간보다 작은)만을 가지고 있다는 것, 또는 사용자에게 의해 또는 미국 특허 제 6,606,411호에 설명된 것과 같은 알고리즘에 의해 영상이 동일한 이벤트에 속한다고 분류된 것을 의미한다. 간단히 요약하면, 영상의 컬렉션은, 영상의 시간 및/또는 날짜 클러스터링(clustering)에 기초한 영상의 컬렉션의 하나 이상의 큰 시간 차이를 판정하고, 이벤트 사이의 하나 이상의 경계를 가지는 것에 기초하여 - 하나 이상의 경계는 하나 이상의 큰 시간 차이에 대응함 -, 다수의 영상을 이벤트로 분류하여, 하나 이상의 이벤트로 분류된다.
- <155> 두 인물의 옷이 정합하고 영상이 동일한 이벤트로부터일 때, 두 인물이 동일한 개인일 가능성은 옷의 고유성에 의존한다. 두 인물 사이에 정합하는 옷이 더 고유하면, 두 인물이 동일한 인물일 가능성은 더 커진다.
- <156> 두 인물이 상이한 이벤트에 속하는 영상으로부터일 때, 옷 사이의 부정합은 인물이 동일한 개인일 가능성에 영향을 미치지 않는다(인물이 옷을 바꾸었을 수 있기 때문에).
- <157> 바람직하게는, 사용자는 사용자 인터페이스를 통해 T_0 값을 조정할 수 있다. 값이 증가함에 따라, 디지털 영상 컬렉션 서브세트(112)는 더 적은 영상 또는 비디오를 포함하나, 디지털 영상 컬렉션 서브세트(112)의 영상 또는 비디오가 관심이 있는 인물을 정말로 포함할 가능성은 증가한다. 이러한 방식으로, 사용자는 검색 결과의 정확성 및 수를 판정할 수 있다.
- <158> 본 발명은 인물을 인식하는 것 이상으로, 도 2와 유사한, 도 15에 도시된 것과 같은 일반 물체 인식 방법으로 일반화될 수 있다. 물체를 포함하는 디지털 영상 컬렉션(102)은 물체 탐색기(408)에 의하여 관심이 있는 물체에 대하여 검색된다. 디지털 영상 컬렉션 서브세트(112)는 인간 사용자에게 의해 재검토되기 위하여 디스플레이(332) 위에 디스플레이된다.
- <159> 관심이 있는 물체에 대한 검색은 사용자에게 의해 다음과 같이 개시된다. 디지털 영상 컬렉션(102)의 영상 또는 비디오는 디스플레이(332) 위에 디스플레이되고 사용자에게 의해 검토된다. 사용자는 하나 이상의 영상에 대한 하나 이상의 라벨을 라벨러(104)로 확립한다. 특징 추출기(106)는 라벨러(104)로부터의 라벨(label(s))과 관련된 디지털 영상 컬렉션으로부터 특징을 추출한다. 특징은 라벨과 관련하여 데이터베이스(114)에 저장된다. 물체 검출기(410)는 선택적으로 라벨 붙이는 것과 특징 추출을 돕기 위하여 사용될 수 있다. 디지털 영상 컬렉션 서브세트(112)가 디스플레이(332) 위에 디스플레이될 때, 사용자는 결과를 재검토할 수 있으며, 디스플레이된 영상에 더 라벨을 붙일 수 있다.
- <160> 라벨러(104)로부터의 라벨은 특정 영상 또는 비디오가 관심이 있는 물체를 포함하는 것을 표시하며 다음 중 하나 이상을 포함한다.
- <161> (1) 영상 또는 비디오 안의 관심이 있는 물체의 이름.
- <162> (2) "인물 A(Person A)" 또는 "인물 B(Person B)"와 같은 식별자 또는 텍스트 열과 같은 관심이 있는 물체와 관련된 식별자(an identifier).
- <163> (3) 영상 또는 비디오 안의 관심이 있는 물체의 위치. 바람직하게는, 관심이 있는 물체의 위치는 관심이 있는 물체를 둘러싸는 박스의 좌표에 의해 특정된다. 사용자는 관심이 있는 물체의 위치를 마우스를 사용하여 예를 들어 눈의 위치를 클릭함으로써 표시할 수 있다. 물체 검출기(410)가 물체를 검출할 때, 물체의 위치는 사용자에게 예를 들어, 디스플레이(332) 위의 물체에 동그라미를 두름으로써 강조될 수 있다. 그리고 나서, 사용자는 강조된 물체에 대한 이름 또는 식별자를 제공할 수 있으며, 그것에 의하여 물체의 위치를 사용자가 제공한 라벨에 관련짓는다.
- <164> (4) 관심이 있는 물체를 포함한다고 생각되는 영상 컬렉션으로부터 영상 또는 비디오에 대한 검색을 하기 위한 표시(an indication).

<165> (5) 영상 안에 있지 않은 관심이 있는 물체의 이름 또는 식별자. 예를 들어, 관심이 있는 물체는 인물, 얼굴, 차, 운송수단 또는 동물일 수 있다.

<166> **구성요소 목록(PARTS LIST)**

<167>	10	영상 포착
<168>	25	하나로 합쳐서 생각된 배경 영역
<169>	40	일반 제어 컴퓨터
<170>	102	디지털 영상 컬렉션
<171>	104	라벨러
<172>	106	특징 추출기
<173>	108	인물 탐색기
<174>	110	인물 검출기
<175>	112	디지털 영상 컬렉션 서브세트
<176>	114	데이터베이스
<177>	202	블록
<178>	204	블록
<179>	206	블록
<180>	207	블록
<181>	208	블록
<182>	210	블록
<183>	212	블록
<184>	214	블록
<185>	220	라벨 붙여진 영상
<186>	222	관심이 있는 인물을 포함한다고 정확하게 생각되는 영상
<187>	224	관심이 있는 인물을 포함한다고 부정확하게 생각되는 영상
<188>	226	라벨
<189>	228	생성된 라벨
<190>	240	로컬 특징 검출기
<191>	242	글로벌 특징 검출기
<192>	244	로컬 특징
<193>	246	글로벌 특징
<194>	250	인물 식별자
<195>	252	관심이 있는 인물의 신원
<196>	254	수정된 특징
<197>	256	인물 분류기
<198>	260	제 1 영상
<199>	262	제 2 영상

<200>	264	인물
<201>	266	인물
<202>	268	인물
<203>	270	얼굴 검출기
<204>	272	포착 시간 분석기
<205>	274	검출된 인물
<206>	282	얼굴 영역
<207>	284	옷 영역
<208>	286	배경 영역
<209>	301	디지털 카메라 폰
<210>	303	플래시
<211>	305	렌즈
<212>	311	CMOS 영상 센서
<213>	312	타이밍 생성기
<214>	314	영상 센서 어레이
<215>	316	A/D 컨버터 회로
<216>	318	DRAM 버퍼 메모리
<217>	320	디지털 프로세서
<218>	322	RAM 메모리
<219>	324	리얼-타임 클락
<220>	325	위치 판정기
<221>	328	펌웨어 메모리
<222>	330	영상/데이터 메모리
<223>	332	컬러 디스플레이
<224>	334	사용자 컨트롤
<225>	340	오디오 코덱
<226>	342	마이크로폰
<227>	344	스피커
<228>	350	무선 모듈
<229>	352	RF 채널
<230>	358	폰 네트워크
<231>	362	독 인터페이스
<232>	364	독/충전기
<233>	370	인터넷
<234>	372	서비스 제공자
<235>	408	물체 탐색기

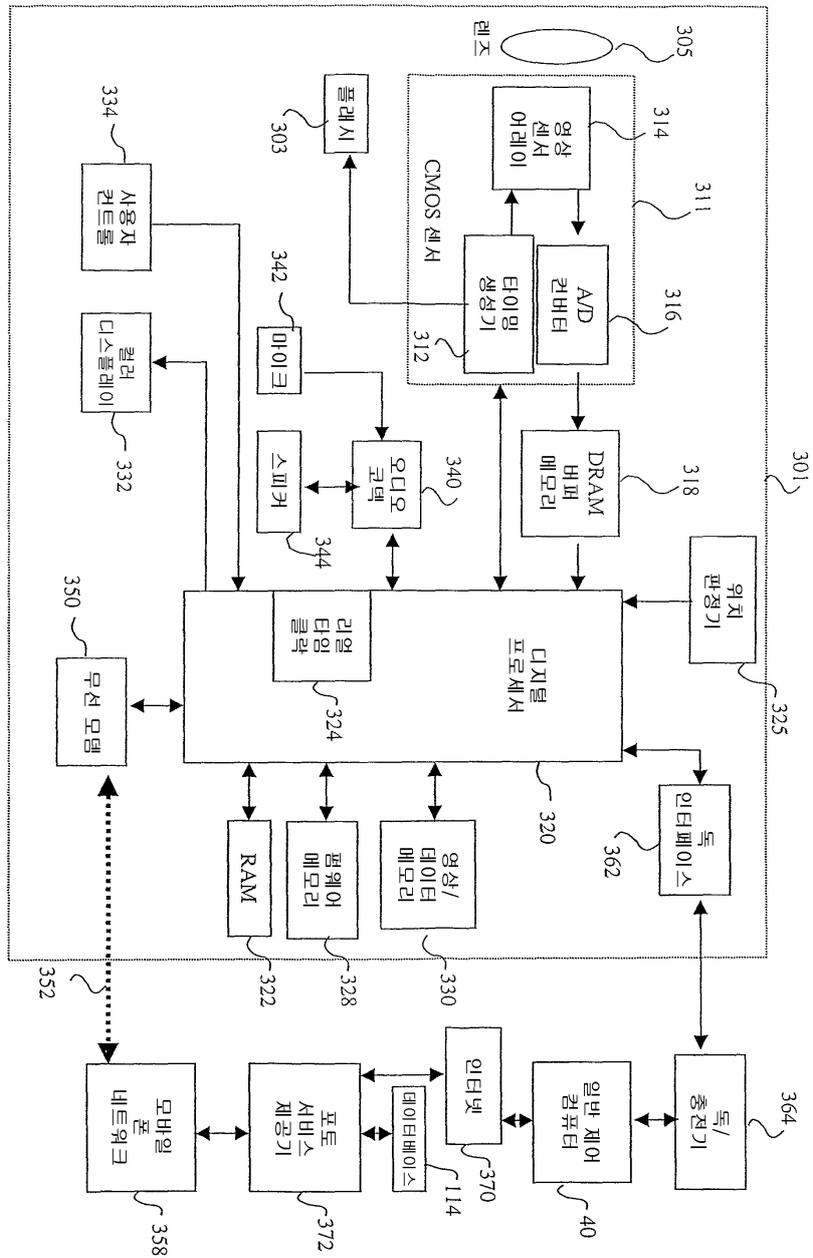
- <236> 410 물체 검출기
- <237> 502 머리카락 영역
- <238> 504 앞머리카락 영역
- <239> 506 안경 영역
- <240> 508 뺨 영역
- <241> 510 긴 머리카락 영역
- <242> 512 턱수염 영역
- <243> 514 콧수염 영역

도면의 간단한 설명

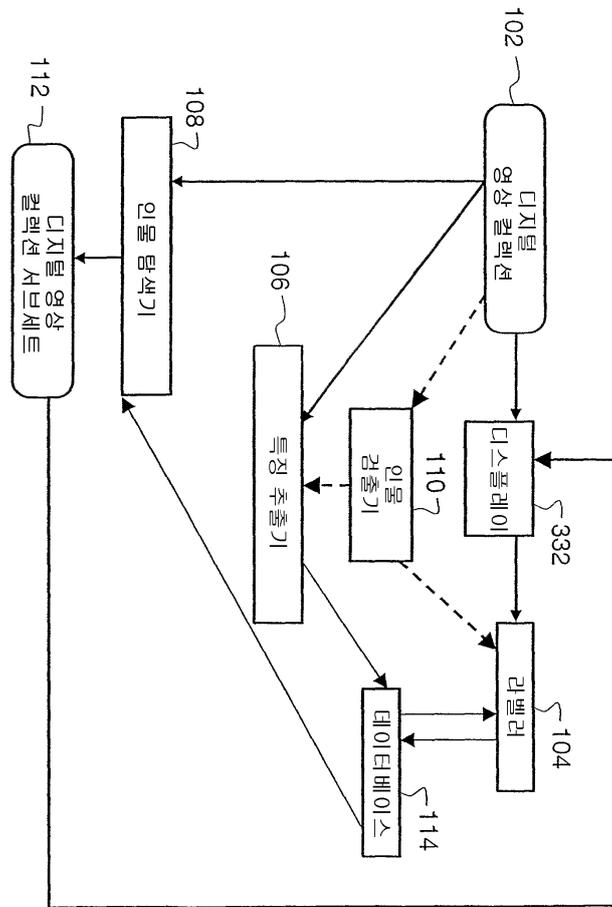
- <15> 본 발명의 요지는 도면에 도시된 실시예를 참고하여 설명될 것이다.
- <16> 도 1은 본 발명을 구현할 수 있는 카메라 폰에 기초한 영상 시스템의 블록도이다.
- <17> 도 2는 디지털 영상 컬렉션에서 관심이 있는 인물을 탐색하기 위한 본 발명의 일 실시예의 순서도이다.
- <18> 도 3은 디지털 영상 컬렉션에서 관심이 있는 인물을 탐색하기 위한 본 발명의 일 실시예의 순서도이다.
- <19> 도 4는 관심이 있는 인물에 대한 검색을 개시하는데 사용되는 영상의 대표적인 세트(set)를 도시한다.
- <20> 도 5는 관심이 있는 인물에 대한 검색의 결과로 사용자에게 디스플레이되는 영상의 대표적인 서브세트(subset)를 도시한다.
- <21> 도 6은 사용자가 관심이 있는 인물을 포함하고 있지 않은 영상을 제거하고 난 후 사용자에게 디스플레이되는 영상의 서브세트를 도시한다.
- <22> 도 7은 디지털 영상 컬렉션에서 관심이 있는 인물을 탐색하기 위한 본 발명의 다른 실시예의 순서도이다.
- <23> 도 8은 영상 및 관련된 라벨을 도시한다.
- <24> 도 9는 관심이 있는 인물에 대한 검색의 결과로 사용자에게 디스플레이되는 영상의 대표적인 서브세트를 도시한다.
- <25> 도 10은 사용자가 관심이 있는 인물을 포함하고 있지 않은 영상을 제거하고 난 후 사용자에게 디스플레이되는 라벨 및 영상의 서브세트를 도시한다.
- <26> 도 11은 도 2로부터의 특징 추출기(feature extractor)의 보다 상세한 보기(more detailed view)를 도시한다.
- <27> 도 12a는 도 2로부터의 인물 검출기(person detector)의 보다 상세한 보기(more detailed view)를 도시한다.
- <28> 도 12b는 영상 포착 시간의 차이 및 하나의 영상에 나타난 인물이 제 2 영상에 또 나타날 확률의 관계를 나타낸 도면이다.
- <29> 도 12c는 영상 포착 시간의 차이의 함수로서 얼굴 크기 비율(facial size ratio)의 관계를 나타낸 도면이다.
- <30> 도 12d는 도 2의 특징 추출기(feature extractor)에 의해 얼굴로부터 추출된 특징 점(feature points)의 표시이다.
- <31> 도 12e는 얼굴 영역(face regions), 옷 영역(clothing regions) 및 배경 영역(background regions)의 표시이다.
- <32> 도 12f는 다양한 얼굴 특징 영역의 표시이다.
- <33> 도 13은 도 2의 인물 탐색기(person finder)의 보다 상세한 보기(more detailed view)를 도시한다.
- <34> 도 14는 15 개의 얼굴에 대한 로컬 특징, 얼굴의 실제의 신원(identity) 및 얼굴의 가능한 신원을 도시한다.
- <35> 도 15는 디지털 영상 컬렉션에서 관심이 있는 물체를 찾기 위한 본 발명의 일 실시예의 순서도이다.

도면

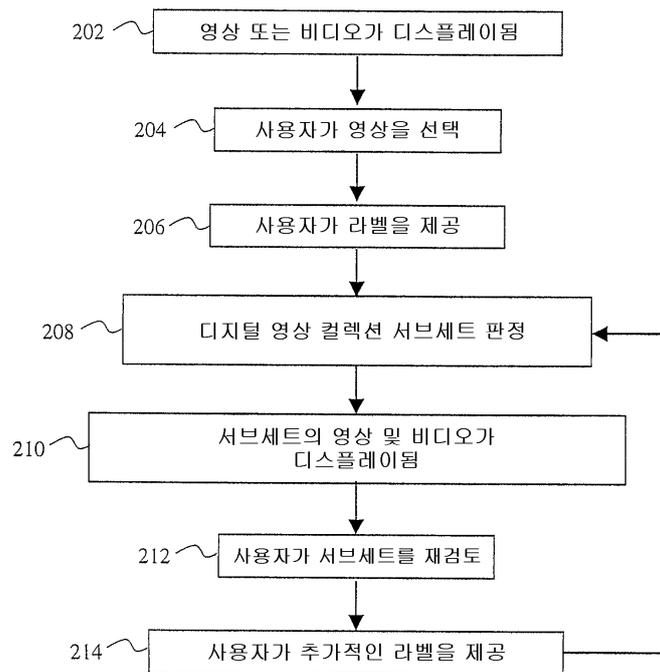
도면1



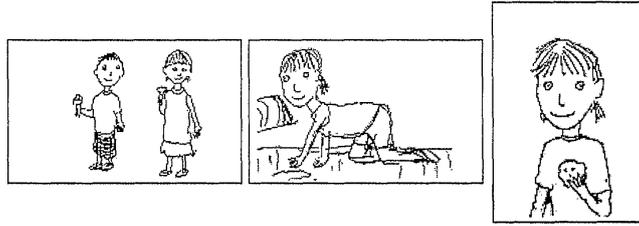
도면2



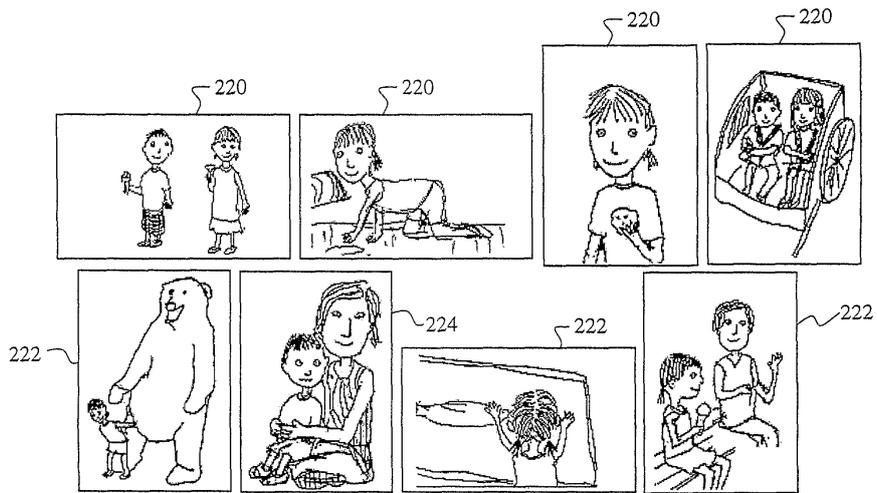
도면3



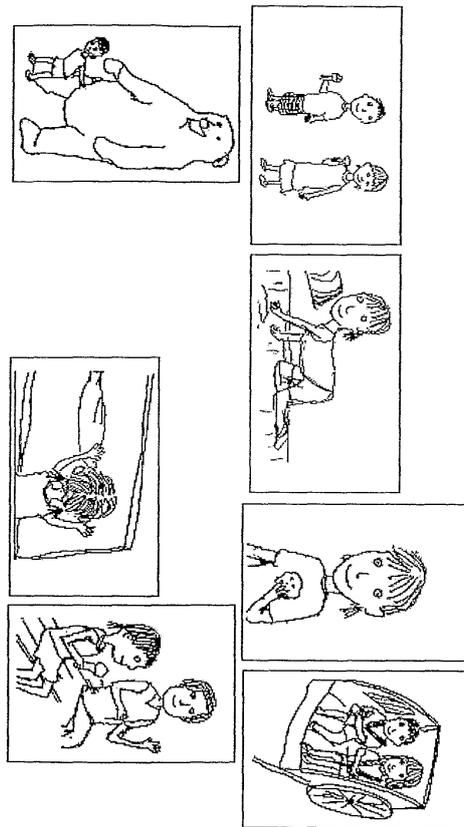
도면4



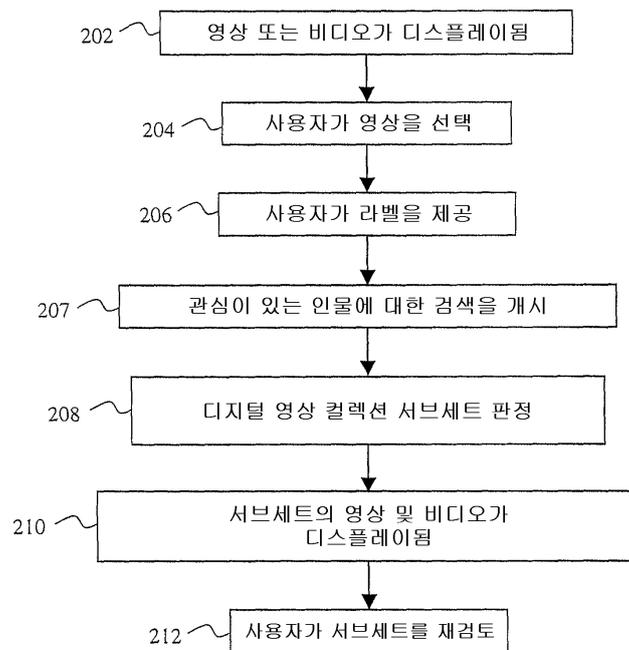
도면5



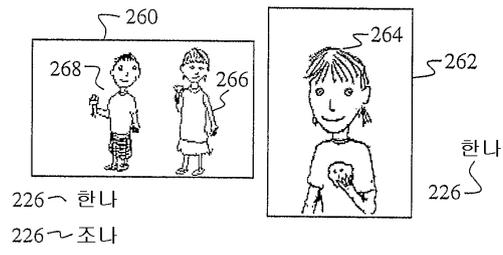
도면6



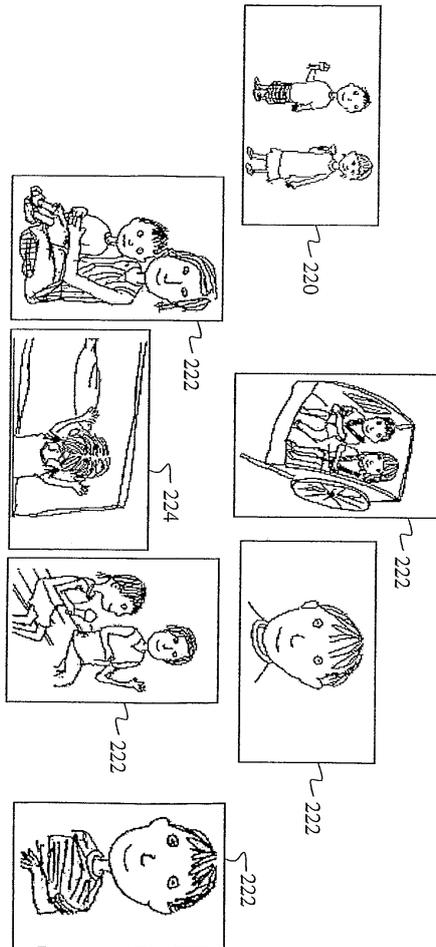
도면7



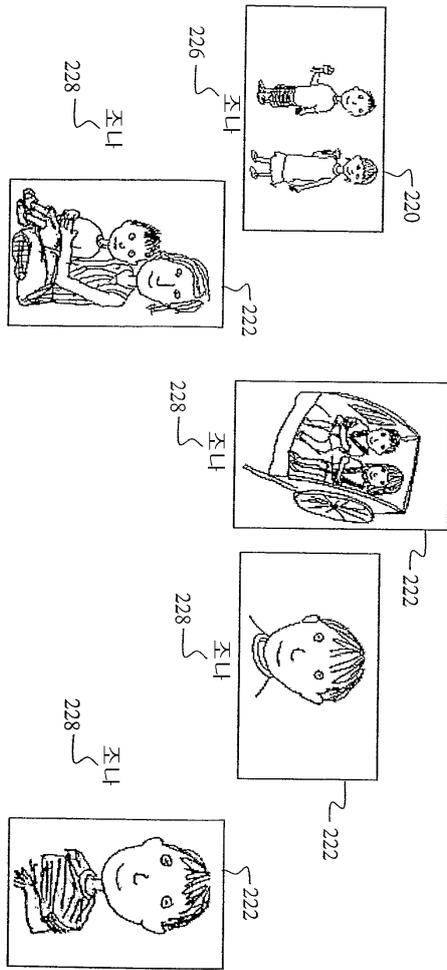
도면8



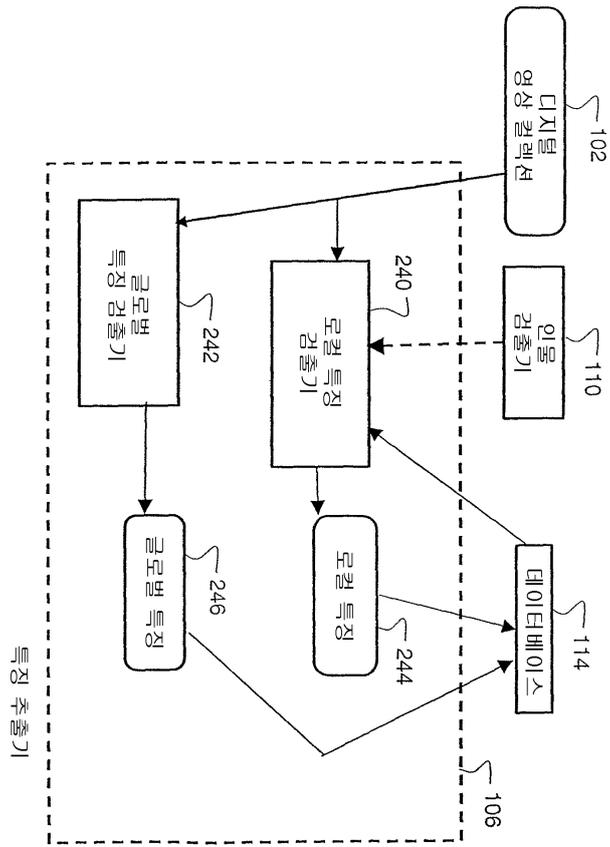
도면9



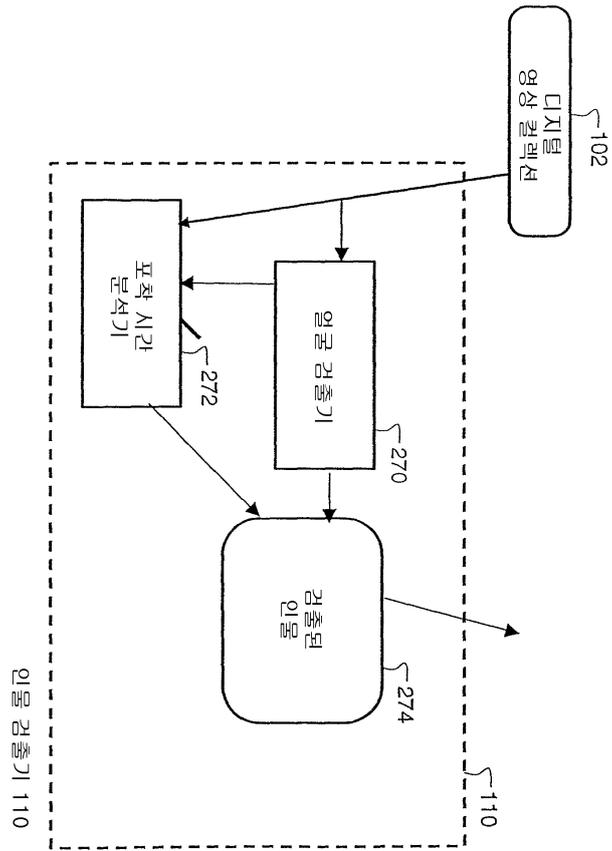
도면10



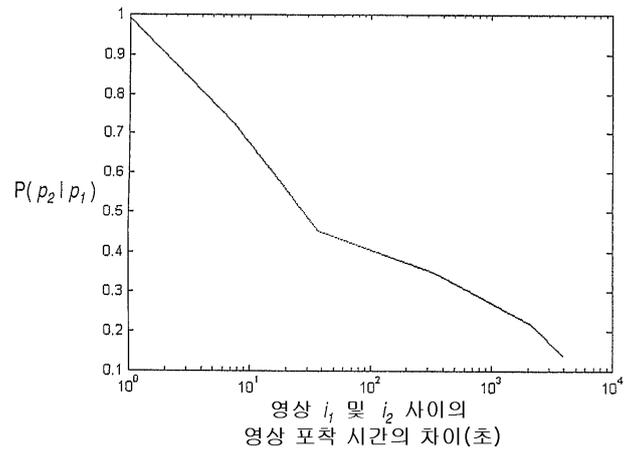
도면11



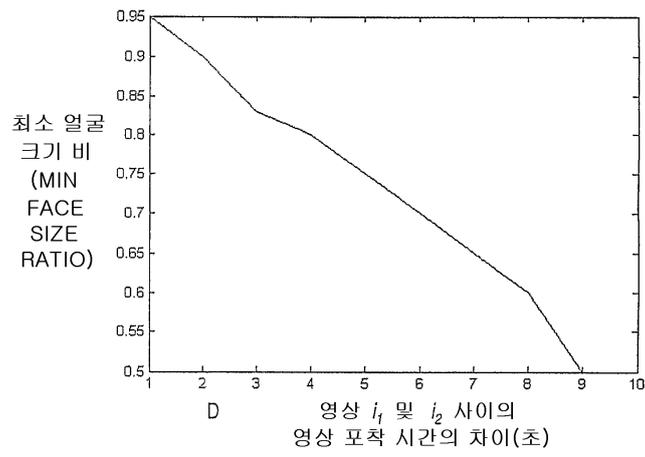
도면12a



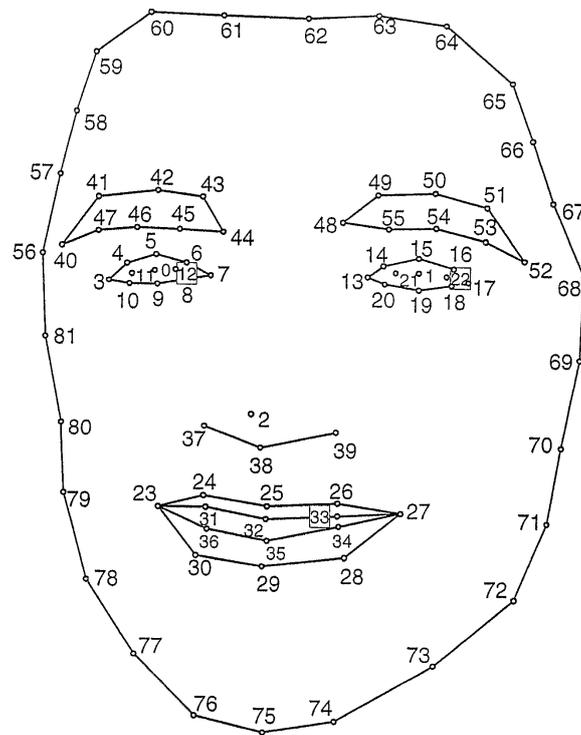
도면12b



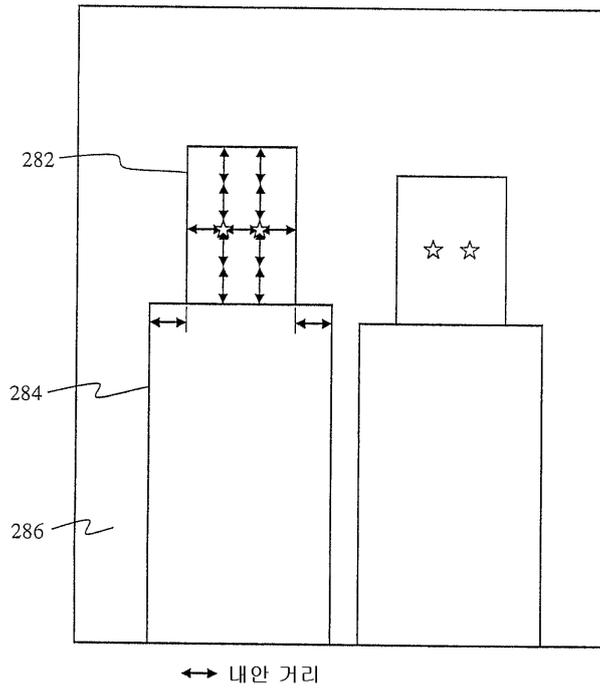
도면12c



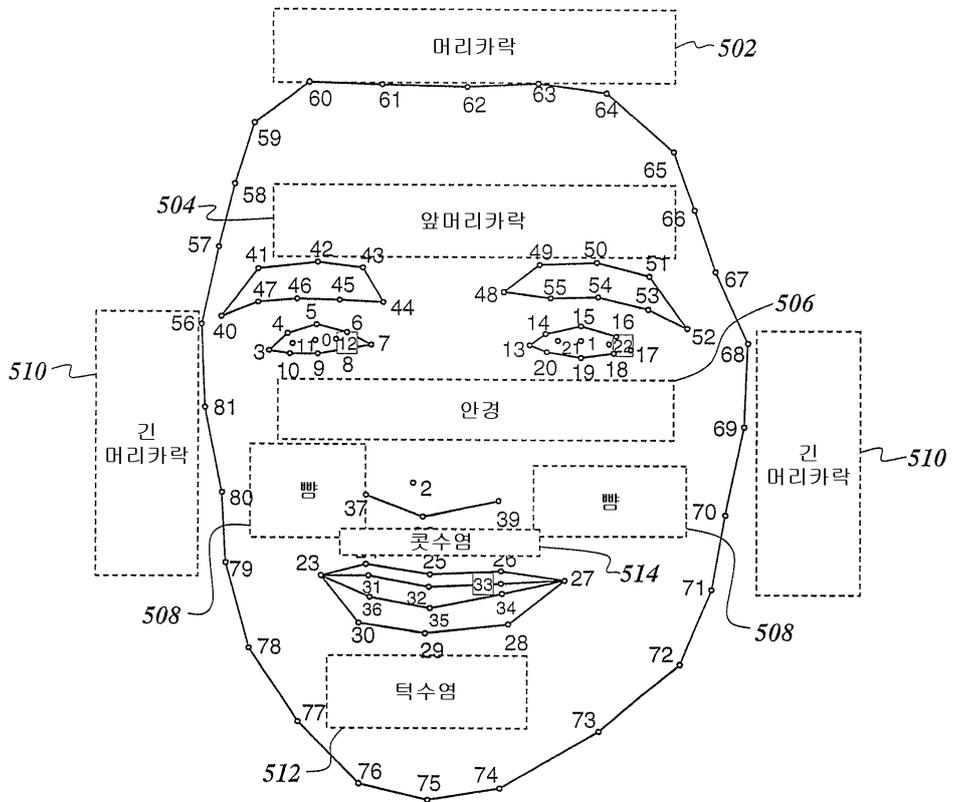
도면12d



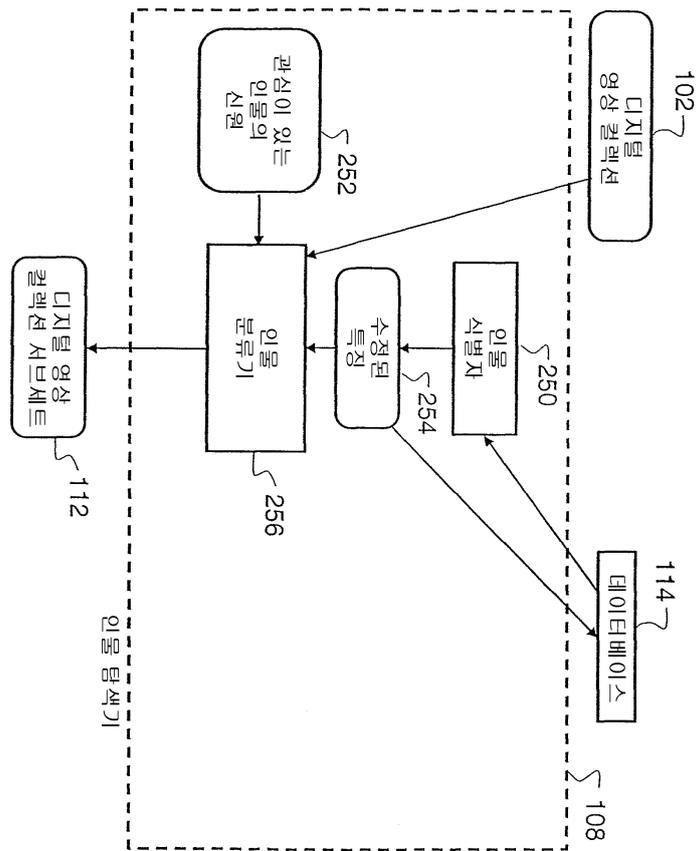
도면12e



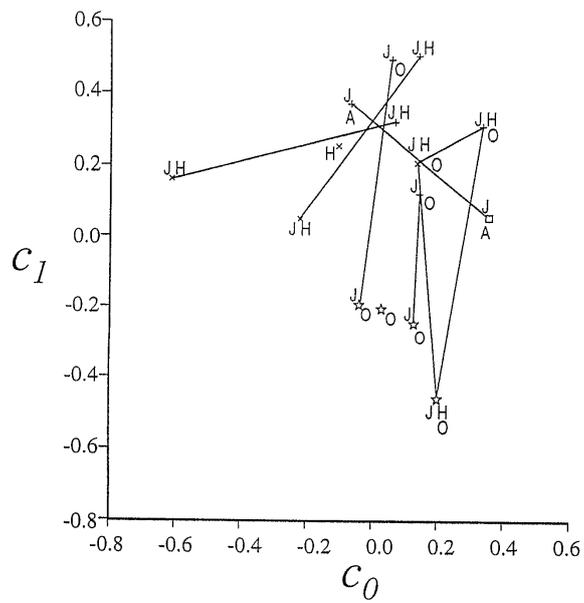
도면12f



도면13



도면14



도면15

