



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2018년12월21일
(11) 등록번호 10-1931837
(24) 등록일자 2018년12월17일

- (51) 국제특허분류(Int. Cl.)
G06T 7/00 (2017.01) *G06K 9/46* (2006.01)
- (21) 출원번호 10-2013-7032354
(22) 출원일자(국제) 2012년06월04일
 심사청구일자 2017년04월27일
(85) 번역문제출일자 2013년12월05일
(65) 공개번호 10-2014-0028064
(43) 공개일자 2014년03월07일
(86) 국제출원번호 PCT/US2012/040788
(87) 국제공개번호 WO 2012/170359
 국제공개일자 2012년12월13일
(30) 우선권주장
 13/299,857 2011년11월18일 미국(US)
 61/493,857 2011년06월06일 미국(US)
(56) 선행기술조사문헌
 US20100215257 A1
 US20100166258 A1

- (73) 특허권자
 마이크로소프트 테크놀로지 라이센싱, 엘엘씨
 미국 워싱턴주 (우편번호 : 98052) 레드몬드 원
 마이크로소프트 웨이
(72) 발명자
 앰브러스 안토니
 미국 워싱턴주 98052-6399 레드몬드 원 마이크로
 소프트 웨이 엘씨에이 - 인터내셔널 페이턴츠 마
 이크로소프트 코포레이션
 리 경석 데이비드
 미국 워싱턴주 98052-6399 레드몬드 원 마이크로
 소프트 웨이 엘씨에이 - 인터내셔널 페이턴츠 마
 이크로소프트 코포레이션
 (뒷면에 계속)
(74) 대리인
 김태홍

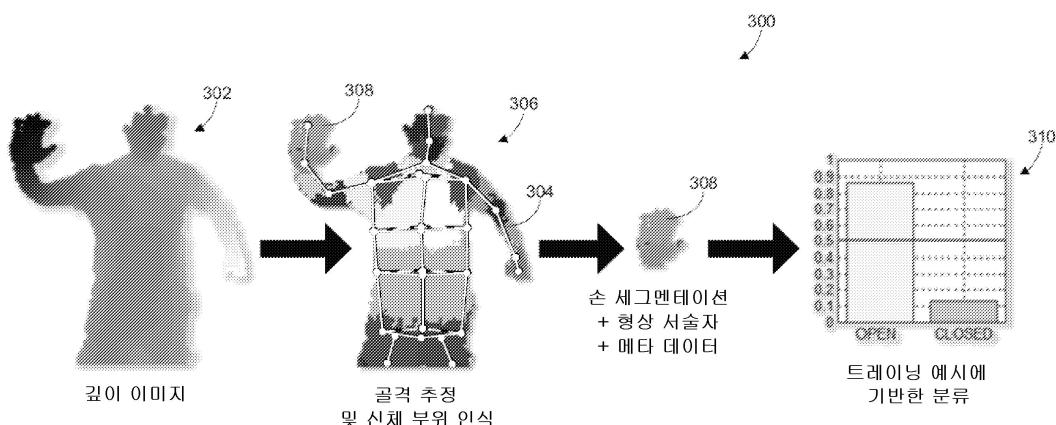
전체 청구항 수 : 총 17 항

심사관 : 신재철

(54) 발명의 명칭 펼친 손 또는 오므린 손의 인식 시스템

(57) 요 약

NUI 시스템의 이미지 센서에서 캡쳐되는, 사용자의 손의 손 모델을 포함하는, 타겟 사용자의 컴퓨터 모델을 생성하기 위한 파이프라인에 관련된 시스템 및 방법이 개시된다. 컴퓨터 모델은 사용자의 한 손 또는 양손의 위치 및 사용자의 한 손 또는 양손이 펼친 상태인지 또는 오므린 상태인지의 최상의 추정치를 나타낸다. 생성된 손모델은 게임 또는 다른 애플리케이션에서 사용되어 사용자의 제스처 및 제어 동작 등을 결정할 수 있다.

대 표 도

(72) 발명자

캠프벨 앤드류

미국 위싱턴주 98052-6399 레드몬드 원 마이크로소프트 웨이 엘씨에이 - 인터내셔널 페이턴츠 마이크로소프트 코포레이션

할리 레이비드

미국 위싱턴주 98052-6399 레드몬드 원 마이크로소프트 웨이 엘씨에이 - 인터내셔널 페이턴츠 마이크로소프트 코포레이션

마운트 브라이언

미국 위싱턴주 98052-6399 레드몬드 원 마이크로소프트 웨이 엘씨에이 - 인터내셔널 페이턴츠 마이크로소프트 코포레이션

로블레스 알버트

미국 위싱턴주 98052-6399 레드몬드 원 마이크로소프트 웨이 엘씨에이 - 인터내셔널 페이턴츠 마이크로소프트 코포레이션

오스본 다니엘

미국 위싱턴주 98052-6399 레드몬드 원 마이크로소프트 웨이 엘씨에이 - 인터내셔널 페이턴츠 마이크로소프트 코포레이션

라이트 손

미국 위싱턴주 98052-6399 레드몬드 원 마이크로소프트 웨이 엘씨에이 - 인터내셔널 페이턴츠 마이크로소프트 코포레이션

샤카시 나힐

미국 위싱턴주 98052-6399 레드몬드 원 마이크로소프트 웨이 엘씨에이 - 인터내셔널 페이턴츠 마이크로소프트 코포레이션

힐 레이브

미국 위싱턴주 98052-6399 레드몬드 원 마이크로소프트 웨이 엘씨에이 - 인터내셔널 페이턴츠 마이크로소프트 코포레이션

맥컬로치 다니엘

미국 위싱턴주 98052-6399 레드몬드 원 마이크로소프트 웨이 엘씨에이 - 인터내셔널 페이턴츠 마이크로소프트 코포레이션

밸란 알렉산드루

미국 위싱턴주 98052-6399 레드몬드 원 마이크로소프트 웨이 엘씨에이 - 인터내셔널 페이턴츠 마이크로소프트 코포레이션

명세서

청구범위

청구항 1

한 개 이상의 손가락들을 포함하는 사용자의 손의 모델을 생성하는 방법에 있어서,

(a) 컴퓨팅 디바이스와 관련된 센서와 인터랙션하는 사용자의 위치를 나타내는 위치 데이터를 상기 컴퓨팅 디바이스에서 수신하는 단계 - 상기 위치 데이터는 상기 사용자의 손을 나타내는 깊이 데이터 및 이미지 데이터 중 적어도 하나를 포함함 - , 및

(b) 상기 위치 데이터를 분석하여 상기 손이 펼쳐진 상태 또는 오므려진 상태에 있는지의 여부를 식별하는 단계를 포함하고, 상기 단계 (b)는,

(b)(1) 상기 단계 (a)에서 캡처된 상기 위치 데이터로부터 깊이 데이터를 분석하여 상기 위치 데이터를 상기 손의 데이터로 세그먼트화하는 단계, 및

(b)(2) 상기 단계 (b)(1)에서 식별된 상기 손의 상기 이미지 데이터에 하나 이상의 필터들을 적용함으로써 피쳐 서술자(feature descriptor)들의 집합(set)을 추출하는 단계 - 상기 하나 이상의 필터들은 상기 손의 경계 밖에 있는 이미지 데이터와 비교하여 상기 손의 이미지 데이터를 분석하여 상기 손의 형상을 포함하는 상기 손의 피쳐들을 구분함 - 를 포함하는 사용자 손 모델 생성 방법.

청구항 2

제 1 항에 있어서, 상기 센서를 통해 명령들(commands)을 수신하는 애플리케이션을 실행하고, 상기 단계 (b)에서 식별된 대로 상기 손이 펼쳐져 있는지 또는 오므려져 있는지의 여부에 기반하여 상기 애플리케이션에서 제어 동작에 영향을 주는 단계를 더 포함하는 사용자 손 모델 생성 방법.

청구항 3

제 1 항에 있어서, 상기 센서를 통해 명령들을 수신하는 게이밍 애플리케이션을 실행하고, 상기 단계 (b)에서 식별된 대로 상기 손이 펼쳐져 있는지 또는 오므려져 있는지의 여부에 기반하여 상기 게이밍 애플리케이션에서 동작에 영향을 주는 단계를 더 포함하는 사용자 손 모델 생성 방법.

청구항 4

제 1 항에 있어서, 상기 단계 (b)(1)은 상기 이미지 데이터로부터 구성되는 중심들(centroids)을 분석하여 손의 최상의 후보(best candidate)를 찾는 단계를 포함하고,

상기 중심들은 손가락 끝 중심들, 손가락 중심들, 손가락 관절 중심들 및 손바닥 중심 중 적어도 하나를 포함하는 것인 사용자 손 모델 생성 방법.

청구항 5

제 4 항에 있어서, 상기 단계 (b)(1)은 상기 손의 최상의 후보를 분석하여 상기 손의 중심의 최상의 후보를 결정하는 단계를 더 포함하는 것인 사용자 손 모델 생성 방법.

청구항 6

제 1 항에 있어서, 상기 단계 (b)(2)는 화소 분류기(pixel classifier)를 적용하는 단계를 포함하고, 상기 화소 분류기를 적용하는 단계는,

상기 손의 형상 서술자의 경계 내의 화소를 선택하는 단계,

각각의 화소 주변에 사전 결정된 크기의 상자를 만드는 단계 - 각각의 상자는 상기 형상 서술자의 평면(plane)에 만들어짐 - ,

상기 이미지 데이터가 전면 포인트와 배경 포인트 사이에서 전환되는 각각의 상자와의 교차점들을 결정하는 단

계, 및

검사된 각각의 화소에 대한 각각의 상자의 교차점들의 분석으로부터 손이 펼쳐져 있는지 또는 오므려져 있는지의 여부를 식별하는 단계를 포함하는 것인 사용자 손 모델 생성 방법.

청구항 7

제 1 항에 있어서, 상기 단계 (b)(2)는 곡률 분석 필터(curvature analysis filter)를 적용하는 단계를 포함하고, 상기 곡률 분석 필터를 적용하는 단계는,

상기 손의 형상 서술자의 경계를 따라 화소들을 선택하는 단계,

선택된 화소를 둘러싸는 복수의 화소들을 검사하고, 둘러싸고 있는 어떤 화소가 또한 상기 형상 서술자의 경계를 따라 있는지를 나타내는 값을 상기 선택된 화소에 부여하는 단계,

정점(peak)들과 골(valley)들을 포함하는 상기 손 주변의 각도들과 윤곽선(contour)들로 상기 값을 변환하는 단계, 및

상기 정점들 중 어떤 정점이 상기 손의 펼친 또는 오므린 손가락들을 나타내는지를 결정하는 단계를 포함하는 것인 사용자 손 모델 생성 방법.

청구항 8

제 1 항에 있어서, 상기 단계 (b)(2)는 내추럴 그랩 필터(natural grab filter)를 적용하는 단계를 포함하고, 상기 내추럴 그랩 필터를 적용하는 단계는 손가락들이 펼쳐진 상태에서 오므려지는 상태로 움직이는 상기 센서를 향하는 상기 손의 움직임을 검출하는 단계를 포함하는 것인 사용자 손 모델 생성 방법.

청구항 9

검출된 손이 펼쳐져 있는지 또는 오므려져 있는지의 여부를 결정하는 시스템에 있어서 - 상기 시스템은 컴퓨팅 장치에 동작가능하게 결합되는 감지 메커니즘(sensing mechanism)을 포함함 - ,

이미지 데이터 및 깊이 데이터 중 적어도 하나를 포함하는 수신되는 데이터로부터 사용자의 골격의 적어도 일부를 인식하는 골격 인식 엔진(skeletal recognition engine),

상기 골격 인식 엔진으로부터 수신되는 깊이 데이터를 분석하여 신체의 하나 이상의 영역들을 상기 사용자의 손을 나타내는 영역으로 세그먼트화하는 이미지 세그멘테이션 엔진(image segmentation engine), 및

한 개 이상의 손가락들을 포함하는 손과 상기 손이 펼쳐져 있는지 또는 오므려져 있는지의 여부를 나타내는 데 이터를 추출하는 서술자 추출 엔진(descriptor extraction engine) - 상기 서술자 추출 엔진은 상기 손을 나타내는 상기 영역 내의 화소들을 분석하도록 복수의 필터들을 적용하고, 상기 복수의 필터들에서 각각의 필터는 상기 손의 위치 및 펼쳐진 또는 오므려진 상태를 결정하며, 상기 서술자 추출 엔진은 각각의 필터의 결과들을 결합하여 상기 손이 펼쳐져 있는지 또는 오므려져 있는지의 여부의 최상의 추정치에 도달함 - 을 포함하는 시스템.

청구항 10

제 9 항에 있어서, 상기 서술자 추출 엔진의 상기 복수의 필터들은 손으로 추정되는 형상 주변의 윤곽선 불연속점(contour discontinuity)들을 측정하는 불연속점 필터(discontinuity filter)를 포함하고, 상기 불연속점 필터는 상기 윤곽선 불연속점들이 사전 결정된 임계치 아래일 때 오므린 손이라고 식별하는 것인 시스템.

청구항 11

제 9 항에 있어서, 상기 서술자 추출 엔진의 상기 복수의 필터들은, 감지 메커니즘을 향하는 손의 움직임과 상기 손의 손가락들이 펼쳐져 있는지 또는 오므려져 있는지의 여부를 검출하는 내추럴 그랩 필터를 포함하는 것인 시스템.

청구항 12

제 11 항에 있어서, 상기 내추럴 그랩 필터는, 상기 손과 동일한 팔의 팔꿈치로 식별된 상기 골격의 포인트로부터 가장 멀리 있는 손의 화소들을 식별하는 것에 부분적으로 의해, 상기 손의 손가락들이 펼쳐져 있는지 또는

오므려져 있는지의 여부를 결정하는 것인 시스템.

청구항 13

제 9 항에 있어서, 상기 서술자 추출 엔진의 상기 복수의 필터들은 상기 손의 손가락들이 펼쳐져 있는지 또는 오므려져 있는지의 여부를 결정하는 하나 이상의 서브-필터들을 더 포함하고, 상기 하나 이상의 서브-필터들 중 제 1 서브-필터는 상기 손의 실루엣이 원에 근사하는지의 여부를 결정하는 원형(circle) 서브-필터이고, 상기 원형 서브-필터는 상기 실루엣이 원에 근사하면 상기 손이 오므려져 있다고 나타내는 것인 시스템.

청구항 14

제 13 항에 있어서, 상기 원형 서브-필터는 상기 손으로 추정된 것의 질량 중심을 식별하고, 상기 질량 중심과 팔꿈치로부터 가장 멀리 있는 것으로 식별된 화소들 사이의 표준 편차를 결정하는 것인 시스템.

청구항 15

제 13 항에 있어서, 상기 하나 이상의 서브-필터들 중 제 2 서브-필터는 깊이 점프(jump)들을 측정하고, 상기 깊이 점프들의 수가 어떤 사전 결정된 값 아래이면 상기 손이 오므려져 있다고 추정하는 것인 시스템.

청구항 16

제 14 항에 있어서, 상기 하나 이상의 서브 필터들 중 제 3 서브 필터는 함께 뻗어 있는 손가락들을 손이 포함하는지의 여부를 결정하는 패들(paddle) 서브-필터이고, 상기 패들 서브-필터는 히스토그램을 사용하여 손가락 끝 화소들을 빈(bin)들로 분류하고, 펼쳐진 손 및 오므려진 손의 공지된 모델들을 사용하여 각각의 빈들 내의 화소들의 수를 검사하여 손이 펼쳐져 있는지 또는 오므려져 있는지의 여부를 추정하는 것인 시스템.

청구항 17

제 16 항에 있어서, 손이 펼쳐져 있는지 또는 오므려져 있는지의 여부를 결정하기 위해 상기 하나 이상의 서브-필터들 각각에 대한 결과 및 신뢰도(confidence)가 결합되는 것인 시스템.

발명의 설명

기술 분야

배경 기술

[0001] 과거에는, 컴퓨터 게임 및 멀티미디어 애플리케이션을 비롯한 컴퓨팅 애플리케이션에서 컨트롤러, 리모콘, 키보드, 마우스 등을 사용하여 사용자들이 게임 캐릭터나 애플리케이션의 기타 양태들을 제어할 수 있었다. 최근 들어, 컴퓨터 게임 및 멀티미디어 애플리케이션은 카메라 및 소프트웨어 제스처 인식 엔진을 이용하여 내츄럴 사용자 인터페이스("NUI")를 제공하기 시작하였다. NUI를 통해, 원 관절 데이터(raw joint data) 및 사용자 제스처를 검출, 해석 및 사용하여 게임 캐릭터나 애플리케이션의 기타 양태들을 제어하게 된다.

[0002] NUI 시스템의 과제 중 하나는 이미지 센서의 시야에 있는 사람을 구분하고, 시야 내에서 손과 손가락을 포함하는 신체 부위의 위치(position)를 정확하게 식별하는 것이다. 팔, 다리, 머리 및 상체를 트래킹하는 루틴이 알려져 있다. 그러나, 사용자 손의 감지하기 힘든 세부 사항과 매우 다양한 위치가 제공되어도, 종래의 시스템은 손가락 및 손 위치를 포함하는 사용자의 신체를 충분히 인식하고 트래킹할 수 없다. 사용자의 손이 펼쳐져 있는지 또는 오므려져 있는지의 검출은 아주 다양한 제어 및/또는 애플리케이션 메트릭(metrics)에서 사용될 수 있다. 사용자 손(들)의 펼친 또는 오므린 상태를 신뢰할만하게 검출할 수 있는 시스템은 유용할 것이다.

발명의 내용

해결하려는 과제

과제의 해결 수단

[0003]

NUI 시스템으로 손과 손가락의 위치 및 손이 펼쳐져 있는지 오므려져 있는지를 포함하는 사용자의 골격 관절을 인식 및 트래킹하는 시스템 및 방법이 본원에 개시된다. 예를 들어, NUI 시스템은 스크린 상의 객체를 선택, 연결(engaging), 잡아 끄는(grabbing and dragging) 등의 이벤트를 트리거하는 데 손과 손가락의 위치의 트래킹을 사용할 수 있다. 손과 손가락의 위치 및 모션을 인식 및 트래킹하는 본 기술에서는 다양한 다른 제스처, 제어 동작 및 애플리케이션을 사용할 수 있다. 사용자의 손의 펼치거나 오므린 상태를 결정함으로써, 사용자와 NUI 또는 다른 시스템의 상호 작용성(interactivity)이 증가할 수 있고, 더욱 간단하고 직관적인 인터페이스가 사용자에게 제공될 수 있다.

[0004]

일례로, 본 개시는 한 개 이상의 손가락을 포함하는 사용자의 손의 모델을 생성하는 방법에 관한 것으로, 이 방법은 (a) 센서와 인터랙션하는 사용자의 위치를 표현하는 위치 데이터를 수신하는 단계 - 위치 데이터는 사용자의 손을 표현하는 깊이 데이터 및 이미지 데이터 중 적어도 하나를 포함함 -, 및 (b) 위치 데이터를 분석하여 손이 펼친 상태 또는 오므린 상태에 있는지를 식별하는 단계를 포함하고, 단계 (b)는 (b)(1) 단계(a)에서 캡쳐된 위치 데이터로부터 깊이 데이터를 분석하여 위치 데이터를 손의 데이터로 세그먼트화하는 단계, 및 (b)(2) 단계 (b)(1)에서 식별된 손의 이미지 데이터에 하나 이상의 필터를 적용함으로써 피쳐 서술자(feature descriptor)의 집합을 추출하는 단계 - 하나 이상의 필터는 손의 경계 밖에 있는 이미지 데이터와 비교하여 손의 이미지 데이터를 분석하여 손의 형상을 포함하는 손의 피쳐를 구분함 -를 포함한다.

[0005]

다른 예시로, 본 개시는 검출된 손이 펼쳐져 있는지 또는 오므려져 있는지를 결정하는 시스템에 관한 것으로, 이 시스템은 컴퓨팅 장치에 동작가능하게 결합되는 감지 메커니즘(sensing mechanism)을 포함하고, 이 시스템은 이미지 데이터 및 깊이 데이터 중 적어도 하나를 포함하는 수신된 데이터로부터 사용자의 골격의 적어도 일부를 인식하는 골격 인식 엔진(skeletal recognition engine), 신체의 하나 이상의 영역을 사용자의 손을 나타내는 영역으로 세그먼트화하는 이미지 세그멘테이션 엔진(image segmentation engine), 및 하나 이상의 손가락을 포함하는 손과 손이 펼쳐져 있는지 또는 오므려져 있는지를 나타내는 데이터를 추출하는 서술자 추출 엔진(descriptor extraction engine) - 서술자 추출 엔진은 손을 나타내는 영역의 화소들을 분석하도록 복수의 필터를 적용하고, 복수의 필터에서 각각의 필터는 손의 위치 및 펼친 또는 오므린 상태를 결정하며, 서술자 추출 엔진은 각각의 필터의 결과를 결합하여 손이 펼쳐져 있거나 또는 오므려져 있는지의 최상의 추정치(best estimate)에 도달함 - 을 포함한다.

[0006]

또 다른 예시에서, 본 개시는 변조된 데이터 신호로 구성되지 않는 컴퓨터-판독 가능 저장 매체에 있어서, 컴퓨터-판독 가능 저장 매체는, 내추럴 사용자 인터페이스를 위한 하나 이상의 손가락을 포함하는 사용자의 손의 모델을 생성하는 방법을 실행하도록 프로세서를 프로그래밍하는 컴퓨터-실행 가능 명령어를 갖고 있으며, 방법은 (a) 내추럴 사용자 인터페이스와 인터랙션하는 사용자의 이미지 데이터를 수신하는 단계, (b) 이미지 데이터를 분석하여 이미지 데이터에서 손을 식별하는 단계, 및 (c) 식별된 손의 이미지 데이터를 사전 결정된 손 위치와 비교하여, 사용자가 다음의 사전 정의된 손 제스처 또는 제어 동작들 - (c)(1) 잡기(grab) 제스처, (c)(2) 놓기(release) 제스처, (c)(3) 손이 가상 객체 주변에서 오므려져 가상 객체를 작동시키는 당기기(pulling) 제스처, (c)(4) 펼친 손을 식별함으로써 동작을 시작하고 계속하기, 및 (c)(5) 오므린 손을 인식하여 동작을 종료하기 - 중 하나를 실행하였는지를 결정하는 단계를 포함한다.

[0007]

본 요약은 아래의 상세한 설명에서 추가적으로 설명되는 일련의 개념을 간략화된 형태로 소개하기 위한 것이다. 본 요약은 특허청구된 대상의 핵심적인 특징 또는 필수적인 특징을 밝히기 위한 것이 아니며, 특허청구된 대상의 범위를 결정하는 데 일조하기 위해 사용되는 것도 아니다. 또한 특허청구된 대상은 본 내용에서 언급된 임의의 또는 모든 단점들을 해결하는 구현에 한정되지 않는다.

도면의 간단한 설명

[0008]

도 1a는 타겟 인식, 분석 및 트래킹 시스템의 예시적인 실시예를 도시한다.

도 1b는 타겟 인식, 분석 및 트래킹 시스템의 예시적인 다른 실시예를 도시한다.

도 1c는 타겟 인식, 분석 및 트래킹 시스템의 예시적인 또 다른 실시예를 도시한다.

도 2는 타겟 인식, 분석 및 트래킹 시스템에서 사용될 수 있는 캡쳐 장치의 예시적인 실시예를 도시한다.

도 3은 인간 타겟을 표현하기 위해 사용되는 예시적인 신체 모델을 도시한다.

도 4는 인간 타겟을 표현하기 위해 사용되는 예시적인 골격 모델의 실질적인 정면도를 도시한다.

도 5는 인간 타겟을 표현하기 위해 사용되는 예시적인 골격 모델의 비스듬한 뷰(skewed view)를 도시한다.

도 6은 본 기술의 실시예에 따른 타겟을 트래킹하기 위한 파이프라인(pipeline)의 순서도를 도시한다.

도 7은 본 개시의 실시예에 따른 사용자의 손의 상태를 결정하는 예시적인 방법을 도시한다.

도 8은 본 개시의 실시예에 따른 이미지 세그멘테이션 엔진(image segmentation engine)의 동작의 순서도이다.

도 9는 본 개시의 실시예에 따른 화소 분류 필터(pixel classification filter)의 동작의 순서도이다.

도 10은 본 개시의 실시예에 따른 화소 분류 필터의 결정 트리(determination tree)이다.

도 11a 및 11b는 본 개시의 실시예에 따른 화소 분류 필터를 사용하는 손가락 끝(fingertip)의 식별을 도시한다.

도 12는 본 개시의 실시예에 따른 화소 분류 필터를 사용하는 손가락 끝 식별을 도시한다.

도 13은 본 개시의 실시예에 따른 화소 분류 필터를 사용하여 식별된 손의 일부를 도시한다.

도 14는 본 개시의 실시예에 따른 화소 분류 필터를 사용하는 손과 손가락의 식별을 도시한다.

도 15는 본 개시의 실시예에 따른 곡률 분석 필터(curvature analysis filter)의 동작의 순서도이다.

도 16은 본 개시의 실시예에 따른 곡률 분석 필터를 사용하는 손과 손가락의 식별을 도시한다.

도 17은 본 개시의 실시예에 따른 깊이 히스토그램(depth histogram)을 사용하는 펼친 손 및 오므린 손 분석을 도시한다.

도 18은 본 개시의 실시예에 따른 통계 필터(statistical filter)들을 사용한 펼친 및 오므린 손 이미지를 도시한다.

도 19는 본 개시의 실시예에 따른 내추럴 그랩 필터(natural grab filter)의 동작의 순서도이다.

도 20은 손 필터에 기반하여 손 위치를 분류하는 관리자 필터(supervisor filter)의 순서도이다.

도 21a는 타겟 인식, 분석 및 트래킹 시스템에서 하나 이상의 제스처를 해석하는 데 사용될 수 있는 컴퓨팅 환경의 예시적인 실시예를 도시한다.

도 21b는 타겟 인식, 분석 및 트래킹 시스템에서 하나 이상의 제스처를 해석하는 데 사용될 수 있는 컴퓨팅 환경의 다른 예시적인 실시예를 도시한다.

발명을 실시하기 위한 구체적인 내용

[0009] 이제 도 1a 내지 도 21b를 참조하여 본 기술의 실시예들이 설명될 것이며, 이는 일반적으로, NUI 시스템의 이미지 센서에 의해 캡쳐되는, 사용자의 손 및/또는 손가락의 손 모델을 포함하는, 타겟 사용자의 컴퓨터 모델을 생성하기 위한 파이프라인에 관한 것이다. 컴퓨터 모델은 캡쳐된 이미지 데이터의 프레임마다 한 번씩 생성될 수 있고, 캡쳐된 프레임에서 사용자의, 포즈를 포함하는 위치의 최상의 추정치를 나타낸다. 게임 또는 기타 애플리케이션은 각각의 프레임에서 생성된 손 모델을 사용하여 사용자 제스처 및 제어 동작 같은 것들을 결정할 수 있다. 손 모델은 추후 모델 결정을 돋기 위해 파이프라인으로 다시 보내진다.

[0010] 먼저 도 1a 내지 도 2를 참조하면, 본 기술을 구현하기 위한 하드웨어는 사용자(18)와 같은 사람 타겟을 인식, 분석 및/또는 트래킹하는 데 사용 가능한 타겟 인식, 분석 및 트래킹 시스템(10)을 포함한다. 타겟 인식, 분석 및 트래킹 시스템(10)의 실시예로 이하에서 설명되는 바와 같은 게임 또는 기타 애플리케이션을 실행하는 컴퓨팅 환경(12)이 포함된다. 컴퓨팅 환경(12)은 게임 및/또는 비-게임 애플리케이션을 실행하는 데 사용될 수 있도록 하드웨어 컴퓨트 및/또는 소프트웨어 컴퓨트를 포함할 수 있다. 일 실시예에서, 컴퓨팅 환경(12)은 본원에 설명된 프로세스를 실행하기 위한 프로세서 판독 가능 저장 장치에 저장된 명령어를 실행 가능한 표준화된 프로세서, 특화된 프로세서, 마이크로 프로세서 등을 비롯한 프로세서를 포함할 수 있다.

[0011] 또한 시스템(10)은 캡쳐 장치에 의해 감지된 하나 이상의 사용자 및/또는 객체에 관한 이미지 및 오디오 데이터를 캡쳐할 수 있는 캡쳐 장치(20)를 포함한다. 실시예에서, 캡쳐 장치(20)는 한 명 이상의 사용자의 신체 및 손

움직임 및/또는 제스처 및 음성과 관련된 정보를 캡쳐하기 위해 사용될 수 있고, 이 정보는 컴퓨팅 환경에서 수신되어 게임 또는 기타 애플리케이션의 양태들을 렌더링, 인터랙션 및/또는 제어하는 데 사용된다. 컴퓨팅 환경(12) 및 캡쳐 장치(20)의 예시들이 아래에서 보다 상세하게 설명된다.

[0012] 타겟 인식, 분석 및 트래킹 시스템(10)의 실시예들이 디스플레이(14)가 있는 오디오/비주얼(A/V) 장치(16)에 연결될 수 있다. 이 장치(16)는 예를 들어, 사용자에게 게임 또는 애플리케이션 비주얼 및/또는 오디오를 제공하는 텔레비전, 전화, 컴퓨터의 모니터, HDTV(high-definition television) 등일 수 있다. 예를 들어, 컴퓨팅 환경(12)은 게임 또는 기타 애플리케이션에 관련된 오디오/비주얼 신호를 제공할 수 있는 그래픽 카드 등의 비디오 어댑터 및/또는 사운드 카드 등의 오디오 어댑터를 포함할 수 있다. A/V 장치(16)는 컴퓨팅 환경(12)으로부터 오디오/비주얼 신호를 수신하여, 이 오디오/비주얼 신호와 관련된 게임 또는 애플리케이션의 비주얼 및/또는 오디오를 사용자(18)에게 제공할 수 있다. 일 실시예에 따르면, 오디오/비주얼 장치(16)는 예를 들어, S-Video 케이블, 동축 케이블, HDMI 케이블, DVI 케이블, VGA 케이블, 컴퓨터 비디오 케이블 등을 통해 컴퓨팅 환경(12)에 연결될 수 있다.

[0013] 실시예에서, 컴퓨팅 환경(12), A/V 장치(16) 및 캡쳐 장치(20)는 협력하여 디스플레이(14) 상에 아바타 또는 온스크린 캐릭터(19)를 렌더링할 수 있다. 예를 들어, 도 1a는 축구 게임 애플리케이션을 플레이하는 사용자(18)를 도시한다. 사용자의 움직임이 트래킹되어 아바타(19)의 움직임을 애니메이션화하는 데 사용된다. 실시예에서, 사용자(18)가 디스플레이(14) 상의 아바타(19)의 움직임과 동작들을 제어하는 움직임 및 제스처를 취할 수 있게끔 아바타(19)는 현실 세계에서의 사용자(18)의 움직임을 흉내낸다.

[0014] 전술한 바와 같이, 골격 매핑 시스템과 같은 모션 추정 루틴은 예컨대, 사용자의 손의 움직임과 같은 감지하기 힘든 제스처를 검출할 수가 없다. 예를 들어, 사용자는 도 1b에 도시된 바와 같이 자신의 손으로 사용자 인터페이스(21)를 스크롤하고 제어함으로써 NUI 시스템(10)과 인터랙션하기를 바랄 수 있다. 또는 사용자는 예컨대, 도 1c의 23 및 25으로 도시된 바와 같이 자신의 손을 펼치고 및/또는 오므리는 등과 같은 다양한 제스처를 취하려고 할 수 있다.

[0015] 따라서, 본원에서 이제부터 설명될 시스템 및 방법은 사용자의 손의 상태를 결정하는 것에 관한 것이다. 예를 들어, 이런 시스템은 스크린의 객체, 예컨대, 객체(27)(도 1c)를 선택, 연결, 또는 잡아 끄는 등의 이벤트를 트리거하는 데 손을 펴고 오므리는 동작을 사용할 수 있다. 한편 이런 동작은 컨트롤러를 사용할 때는 버튼을 누르는 것에 대응할 것이다. 이와 같이 개선된 컨트롤러-없는 인터랙션은, 직관적이지 않거나 번거로울 수 있는 손 흔들기 또는 호버링(hovering)에 기반한 접근 방법의 대안으로 사용될 수 있다. 손의 모션을 인식 및 트래킹하는 본 기술에서는 다양한 다른 제스처, 제어 동작 및 애플리케이션을 사용할 수 있으며, 그 중 일부가 아래에서 추가적으로 상세하게 설명된다. 아래에서 설명될 바와 같이 사용자의 손의 상태를 결정함으로써, 사용자와 시스템의 상호 작용성이 증가하고, 보다 간단하고 직관적인 인터페이스가 사용자에게 제공될 수 있다.

[0016] 도 1a 및 도 1b는 바닥, 의자 및 식물과 같은 정적인 배경 객체들(23)을 포함한다. 이들은 캡쳐 장치(20)에서 캡쳐된 시야(FOV) 내의 객체들이며, 프레임마다 바뀌지 않는다. 도시된 바닥, 의자 및 식물에 더해, 정적 객체들은 캡쳐 장치(20)의 이미지 카메라에 의해 얻어진 임의의 객체들일 수 있다. 장면 내의 추가적인 정적 객체들로 벽, 천장, 창문, 문, 벽 장식 등을 포함할 수 있다.

[0017] 시스템(10)과 그 컴퓨팅 장치들의 적절한 예시들은 다음의 동시 계류 중인 특허 출원들: "Environment And/Or Target Segmentation"이라는 발명의 명칭으로 2009년 5월 29일에 출원된 미국 특허 출원 제 12/475,094호, "Auto Generating a Visual Representation"이라는 발명의 명칭으로 2009년 7월 29일에 출원된 미국 특허 출원 제 12/511,850호, "Gesture Tool"이라는 발명의 명칭으로 2009년 5월 29일에 출원된 미국 특허 출원 제 12/474,655호, "Pose Tracking Pipeline"이라는 발명의 명칭으로 2009년 10월 21일에 출원된 미국 특허 출원 제 12/603,437호, "Device for Identifying and Tracking Multiple Humans Over Time"이라는 발명의 명칭으로 2009년 5월 29일에 출원된 미국 특허 출원 제 12/475,308호, "Human Tracking System"이라는 발명의 명칭으로 2009년 10월 7일에 출원된 미국 특허 출원 제 12/575,388호, "Gesture Recognizer System Architecture"라는 발명의 명칭으로 2009년 4월 13일에 출원된 미국 특허 출원 제 12/422,661호, 및 "Standard Gestures"라는 발명의 명칭으로 2009년 2월 23일에 출원된 미국 특허 출원 제 12/391,150호에서 발견되며, 이들 모두는 특히 참조로서 본원에 통합된다.

[0018] 도 2는 타겟 인식, 분석 및 트래킹 시스템(10)에서 사용 가능한 캡쳐 장치(20)의 예시적인 실시예를 도시한다. 예시적인 실시예에서, 캡쳐 장치(20)는 TOF(time-of-flight), 구조광(structured light), 스테레오 이미지 등을 포함하는 적절한 임의 기술을 통해, 깊이 값을 포함하는 깊이 이미지 등 깊이 정보가 있는 비디오를 캡쳐하

도록 구성될 수 있다. 일 실시예에 따르면, 캡쳐 장치(20)는 "Z 레이어들", 또는 깊이 카메라로부터 시선을 따라 연장되는 Z 축에 수직인 레이어들로 깊이 정보를 체계화시킬 수 있다. X와 Y 축은 Z 축에 직각인 것으로 정의될 수 있다. Y 축은 수직이고, X 축은 수평일 수 있다. 함께, X, Y 및 Z 축은 캡쳐 장치(20)에서 캡쳐되는 3-D 현실 세계 공간을 정의한다.

[0019] 도 2에 도시된 바와 같이, 캡쳐 장치(20)는 이미지 카메라 컴포넌트(22)를 포함할 수 있다. 일 실시예에 따르면, 이미지 카메라 컴포넌트(22)는 장면의 깊이 이미지를 캡쳐하는 깊이 카메라일 수 있다. 깊이 이미지는 캡쳐된 장면의 2-D(two-dimensional) 화소 영역을 포함할 수 있고, 2-D 화소 영역의 각 화소는 카메라로부터 캡쳐된 장면에서 객체의 길이 또는 거리 등의 깊이 값을 센티미터, 밀리미터 등으로 나타낼 수 있다.

[0020] 도 2에 도시된 바와 같이, 예시적인 실시예에 따르면, 이미지 카메라 컴포넌트(22)는 장면의 깊이 이미지를 캡쳐하기 위해 사용되는 적외선 컴포넌트(IR light component, 24), 3-D(three-dimensional) 카메라(26) 및 RGB 카메라(28)를 포함할 수 있다. 예를 들면, TOF 분석법에서는, 캡쳐 장치(20)의 적외선 컴포넌트(24)가 장면으로 적외선을 내보내고, 센서들(도시되지 않음)을 사용하여 3-D 카메라(26) 및/또는 RGB 카메라(28)를 사용한 장면의 하나 이상의 타겟 및 객체들의 표면으로부터 후방 산란되는 빛을 검출한다.

[0021] 몇몇 실시예에서는, 펄스 적외선(pulsed infrared light)을 사용하여, 나가는 광 펄스와 이에 대응하여 들어오는 광 펄스 간의 시간을 측정하고, 이를 사용하여 캡쳐 장치(20)로부터 장면의 타겟 또는 객체들의 특정 지점까지의 물리적인 거리를 결정할 수 있다. 또한, 또 다른 실시예에서는, 나가는 광파의 위상과 들어오는 광파의 위상을 비교하여 위상 변위(phase shift)를 결정할 수 있다. 이후, 이 위상 변위를 사용하여 캡쳐 장치(20)로부터 타겟 또는 객체의 특정 지점까지의 물리적인 거리를 결정할 수 있다.

[0022] 다른 실시예에 따르면, TOF 분석법을 사용하여 셔터 광 펄스 이미징(shuttered light pulse imaging)을 포함하는 다양한 기술을 통해 시간에 따른 반사광의 강도를 분석함으로써 캡쳐 장치(20)로부터 타겟 또는 객체의 특정 지점까지의 물리적인 거리를 간접적으로 결정할 수 있다.

[0023] 또 다른 실시예에서, 캡쳐 장치(20)는 구조광을 사용하여 깊이 정보를 캡쳐할 수 있다. 이러한 분석법에서는, 패턴화된 광(즉, 그리드 패턴이나 스트라이프 패턴과 같이 공지된 패턴으로 디스플레이되는 광)이 적외선 컴포넌트(24)를 통해 장면에 투사될 수 있다. 그 장면의 하나 이상의 타겟 또는 객체의 표면에 부딪치면, 패턴이 그에 따라 변형되게 된다. 이러한 패턴의 변형이 예컨대, 3-D 카메라(26) 및/또는 RGB 카메라(28)에 의해 캡쳐된 후에 분석되어, 캡쳐 장치(20)로부터 타겟 또는 객체의 특정 지점까지의 물리적인 거리를 결정할 수 있다.

[0024] 또 다른 실시예에 따르면, 캡쳐 장치(20)는, 깊이 정보의 생성을 위해 분석되는 비주얼 스테레오 데이터를 얻기 위해 각기 다른 각도에서 장면을 볼 수 있는 물리적으로 분리된 둘 이상의 카메라들을 포함할 수 있다. 또 다른 실시예에서, 캡쳐 장치(20)는 포인트 클라우드 데이터(point cloud data) 및 타겟 디지털화 기법을 사용하여 사용자의 피처들을 검출할 수 있다. 추가적인 실시예에서 예컨대, x, y 및 z 축을 검출할 수 있는 초음파 시스템과 같은 다른 센서 시스템이 사용될 수 있다.

[0025] 캡쳐 장치(20)는 마이크로폰(30)을 더 포함할 수 있다. 마이크로폰(30)은 소리를 수신하여 이를 전기 신호로 변환하는 변환기(transducer) 또는 센서를 포함할 수 있다. 일 실시예에 따르면, 마이크로폰(30)은 타겟 인식, 분석 및 트래킹 시스템(10)의 캡쳐 장치(20)와 컴퓨팅 환경(12) 사이의 피드백을 감소시키기 위해 사용될 수 있다. 게다가, 컴퓨팅 환경(12)에서 실행 가능한 게임 애플리케이션, 비게임 애플리케이션 등의 애플리케이션들을 제어하기 위해 사용자로부터 제공되는 오디오 신호를 수신하기 위해 마이크로폰(30)이 사용될 수 있다.

[0026] 일 실시예에서, 캡쳐 장치(20)는 이미지 카메라 컴포넌트(22)와 통신 가능한 논리 프로세서(32)를 더 포함할 수 있다. 프로세서(32)는 깊이 이미지를 수신하고, 그 깊이 이미지에 적절한 타겟의 포함 여부를 결정하고, 적절한 타겟을 타겟의 골격 형상 또는 모델로 변환하는 명령어들을 포함하는 명령어들 또는 기타 임의의 적절한 명령어들을 실행하는 표준화된 프로세서, 특화된 프로세서, 마이크로 프로세서를 포함할 수 있다.

[0027] 캡쳐 장치(20)는, 프로세서(32)로 실행 가능한 명령어들, 3-D 카메라 또는 RGB 카메라로 캡쳐된 이미지 또는 이미지 프레임, 또는 기타 임의의 적절한 정보, 이미지 등을 저장하는 메모리 컴포넌트(34)를 더 포함할 수 있다. 일 실시예에 따르면, 메모리 컴포넌트(34)는 RAM(random access memory), ROM(read only memory), 캐시(cache), 플래시 메모리, 하드 디스크 또는 기타 적절한 저장 컴포넌트를 포함한다. 도 2에 도시된 바와 같이, 일 실시예에서, 메모리 컴포넌트(34)는 이미지 카메라 컴포넌트(22)와 프로세서(32)와 통신하는 별개의 컴포넌트일 수 있다. 또 다른 실시예에 따르면, 메모리 컴포넌트(34)가 프로세서(32) 및/또는 이미지 카메라 컴포넌트(22)에 통합될 수도 있다.

- [0028] 도 2에 도시된 바와 같이, 캡쳐 장치(20)는 통신 링크(36)를 통해 컴퓨팅 환경(12)과 통신할 수 있다. 통신 링크(36)는 USB 연결, 파이어와이어(Firewire) 연결, 이더넷 케이블 연결 등을 포함하는 유선 연결 및/또는 무선 802.11b, g, a 또는 n 연결 등의 무선 연결일 수 있다. 일 실시예에 따르면, 컴퓨팅 환경(12)은 장면을 언제 캡쳐할지를 결정하는 데에 사용되는 클러스터링 링크(36)를 통해 캡쳐 장치(20)로 제공할 수 있다.
- [0029] 또한, 캡쳐 장치(20)는 예를 들어, 3-D 카메라(26) 및/또는 RGB 카메라(28)에서 캡쳐되는 깊이 정보 및 이미지들을 제공할 수 있다. 이를 장치를 사용하여, 통신 링크(36)를 통해 컴퓨팅 환경(12)에 제공된 결과 데이터로 본 기술에 따른 부분 골격 모델이 개발될 수 있다.
- [0030] 컴퓨팅 환경(12)은 아래에서 설명될 바와 같이 제스처를 인식하는 제스처 인식 엔진(190)을 더 포함할 수 있다. 본 시스템에 따르면, 컴퓨팅 환경(12)은 골격 인식 엔진(192), 이미지 세그멘테이션 엔진(194), 서술자 추출(descriptor extraction) 엔진(196), 및 분류기(classifier) 엔진(198)을 더 포함할 수 있다. 이를 각각의 소프트웨어 엔진은 이하에서 더욱 상세하게 설명된다.
- [0031] 도 3은 골격 인식 엔진(192)에서 생성되는 예시적인 신체 모델(70)의 비제한적인 비주얼 형상(visual representation)을 도시한다. 신체 모델(70)은 모델화된 타겟(modeled target)(예컨대, 도 1a와 도 1b의 사용자(18))의 기계 형상(machine representation)이다. 신체 모델은 게임 또는 기타 애플리케이션/운영 체제의 언어로 모델화된 타겟을 전체적으로 정의하는 변수 집합을 포함하는 하나 이상의 데이터 구조를 포함할 수 있다.
- [0032] 타겟의 모델은 본 개시의 범위를 벗어나지 않고 다양하게 구성될 수 있다. 몇 가지 예로, 모델은 고정(rigid) 및/또는 변형 가능(deformable) 형태 또는 신체 부위를 포함하는 삼차원 모델로 타겟을 표현하는 하나 이상의 데이터 구조를 포함할 수 있다. 각각의 신체 부위는 수학적 기본형(mathematical primitive)으로 나타낼 수 있고, 수학적 기본형의 예로 구(sphere), 비등방 축척의 구(anisotropically-scaled sphere), 원기둥, 비등방성 원기둥, 부드러운 원기둥(smooth cylinder), 상자, 경사진 상자(beveled box), 프리즘 등을 포함하지만 이에 제한되지는 않는다.
- [0033] 예를 들어, 도 3의 신체 모델(70)은 신체 부위(bp1에서 bp14)를 포함하고, 이를 각각은 모델화된 타겟의 다른 부위를 표현한다. 각각의 신체 부위는 삼차원 형태이다. 예를 들어, bp3는 모델화된 타겟의 원손을 표현하는 직각 프리즘이고, bp5는 모델화된 타겟의 원쪽 상부 팔을 표현하는 팔각형의 프리즘이다. 신체 모델은 임의의 수의 신체 부위를 포함할 수 있다는 점에서 신체 모델(70)은 그 예시가 되고, 신체 부위 각각은 모델화된 타겟의 대응 부위의 임의의 기계-관독 가능 형상일 수 있다.
- [0034] 둘 이상의 신체 부위를 포함하는 모델은 또한, 하나 이상의 관절을 포함할 수 있다. 하나 이상의 신체 부위는 각각의 관절을 통해 하나 이상의 다른 신체 부위에 대해 움직일 수 있게 된다. 예를 들어, 인간 타겟을 표현하는 모델은 복수의 고정 및/또는 변형 가능 신체 부위를 포함할 수 있고, 여기서 일부 신체 부위는 인간 타겟의 해부학상 대응 신체 부위를 표현할 수 있다. 또한, 모델의 각각의 신체 부위는 하나 이상의 구조 부재(structural member)(즉, "뼈" 또는 골격 부위)와 인접 뼈들의 교차점에 위치한 관절을 포함할 수 있다. 일부 뼈는 인간 타겟의 해부학상 뼈에 대응할 수 있고, 및/또는 일부 뼈는 인간 타겟의 해부학상 대응되는 뼈가 없을 수 있음을 이해할 것이다.
- [0035] 뼈와 관절은 골격 모델을 전체적으로 형성하고, 이는 신체 모델의 구성 요소가 될 수 있다. 몇몇 실시예에서, 도 3의 모델(70) 등 다른 유형의 모델을 대신하여 골격 모델을 사용할 수 있다. 골격 모델은 각각의 신체 부위에 대해 하나 이상의 골격 부재와 인접 골격 부재들 사이의 관절을 포함할 수 있다. 예시적인 골격 모델(80)과 예시적인 골격 모델(82)이 각각 도 4와 5에 도시된다. 도 4는 관절(j1에서 j33)을 포함하는, 정면에서 본 골격 모델(80)을 도시한다. 도 5는 역시 관절(j1에서 j33)을 포함하는 비스듬한 뷰로 보여지는 골격 모델(82)을 도시한다. 골격 모델은 본 개시의 사상을 벗어나지 않고 더 많거나 더 적은 관절을 포함할 수 있다. 이제부터 설명되는 본 시스템의 추가적인 실시예는 31 개의 관절을 갖고 있는 골격 모델을 사용하여 실시된다.
- [0036] 전술한 신체 부위 모델 및 골격 모델은 모델화된 타겟의 기계 형상으로써 사용 가능한 모델 유형의 비제한적인 예시들이다. 다른 모델도 본 개시의 범위에 포함된다. 예를 들어, 일부 모델은 다각형 메쉬(polygonal mesh), 패치(patch), NURBS(non-uniform rational B-spline), 세분 표면(subdivision surface), 또는 기타 고차원 표면을 포함할 수 있다. 모델은, 모델화된 타겟의 옷, 머리, 및/또는 기타 양태를 더욱 정확하게 표현하기 위해 표면의 질감 및/또는 다른 정보도 포함할 수 있다. 모델은 현재 포즈, 한 개 이상의 과거 포즈, 및/또는 모델 물리학에 관련된 정보를 선택적으로 포함할 수 있다. 포즈를 취할 수 있는 여러 다른 모델들이 본원에서 설명된 타겟 인식, 분석 및 트래킹 시스템과 호환 가능함을 이해할 것이다.

- [0037] 캡쳐 장치(20)의 FOV 내에서 한 명 이상의 사용자의 골격 모델을 생성하는 소프트웨어 파이프라인이 공지되어 있다. 이런 시스템은 예컨대, "System For Fast, Probabilistic Skeletal Tracking"라는 발명의 명칭으로 2010년 9월 7일에 출원된 미국 특허 출원 제 12/876,418호에 개시되어 있으며, 이는 그 전체가 참조로서 본원에 통합된다. 예컨대, 사용자가 캡쳐 장치(20)에 충분히 가깝게 있고, 사용자의 적어도 한 쪽 손이 다른 배경 노이즈와 구분 가능한 특정 조건 하에서, 소프트웨어 파이프라인은 FOV 내의 한 명 이상의 사용자의 손 및/또는 손가락에 대한 손 모델을 생성할 수 있다.
- [0038] 도 6은 사용자의 손 및/또는 손가락을 인식 및 트래킹하기 위한 소프트웨어 파이프라인의 순서도이다. 단계(200)에서, 파이프라인은 캡쳐 장치(20)로부터 깊이 이미지를 수신한다. 사용자의 일부의 깊이 이미지는 도 7의 302에 도시된다. 깊이 이미지의 각각의 화소는 예컨대, 도 7에 그레이-스케일 그래디언트(gray-scale gradient)에 의해 도시된 바와 같이, 깊이 정보를 포함한다. 예를 들어, 302에서, 왼손의 더 어두운 영역으로 표시된 바와 같이, 사용자의 왼손이 캡쳐 장치(20)에 더 가깝다. 캡쳐 장치 또는 깊이 카메라는 관찰 장면 내에서의 사용자의 이미지를 캡쳐한다. 이하에 설명되는 바와 같이, 사용자의 깊이 이미지를 사용하여 사용자 영역의 거리 정보, 사용자의 크기(scale) 정보, 곡률, 및 사용자의 골격 정보를 결정할 수 있다.
- [0039] 단계(204)에서, 파이프라인의 골격 인식 엔진(192)은 전술한 바와 같이 사용자의 골격 모델을 추정하여, 단계(200)에서 얻은 깊이 이미지로부터 가상 골격을 얻는다. 예를 들어, 도 7에서, 사용자의 302에서 도시된 깊이 이미지로부터 추정된 가상 골격(304)이 도시된다.
- [0040] 단계(208)에서, 파이프라인은, 파이프라인의 이미지 세그멘테이션 엔진(194)을 통해 사용자의 한쪽 손 또는 양 손을 세그먼트화한다. 몇몇 예시에서, 이미지 세그멘테이션 엔진(194)은 양손은 물론 신체의 하나 이상의 부분을 추가적으로 세그먼트화할 수 있다. 사용자의 손을 세그먼트화 하는 단계는 손에 대응하는 깊이 이미지의 영역을 식별하는 단계를 포함하고, 이 식별 단계는 단계(204)에서 얻어지는 골격 정보에 적어도 부분적으로 기반한다. 도 7은, 다르게 음영 처리된 영역들로 표시된 바와 같이, 추정된 골격(304)에 기반하여 사용자의 깊이 이미지를 다른 영역들(306)로 세그먼트화한 예시를 도시한다. 도 7은 사용자가 들어올린 오른 손에 대응하는 국부적인 손 영역(308)을 도시한다.
- [0041] 손과 신체 영역은 다양한 방식으로 세그먼트화되거나 또는 국부화될(localized) 수 있고, 전술한 골격 추정에서 식별된, 선택 관절에 기반할 수 있다. 일례로써, 깊이 이미지의 손 검출 및 국부화는 추정된 골격에서의 추정된 손목 및/또는 손가락 끝 관절에 기반할 수 있다. 예를 들어, 몇몇 실시예에서, 깊이 이미지의 손 세그멘테이션은, 손 관절 주변의 깊이 이미지의 지형 탐색(topographical search)을 사용하여 손가락 끝의 후보로써 깊이 이미지에서 근처 국부 극값(local extrema)을 찾음으로써 실행될 수 있다. 그 다음에 이미지 세그멘테이션 엔진(194)은 추정된 골격에서 결정된 신체 사이즈 스케일링 팩터(scaling factor)뿐만 아니라 경계 구분을 위해 깊이 불연속점(depth discontinuities)까지도 고려하여 손의 나머지를 세그먼트화한다.
- [0042] 다른 예를 들면, 사용자 손에 대응하는 깊이 이미지의 영역을 식별하기 위해 플러드-필(flood- fill) 접근 방법을 이용할 수 있다. 플러드-필 접근 방법에서, 시작점과 시작 방향으로 깊이 이미지가 탐색될 수 있고, 예를 들어, 시작점은 손목 관절이고, 시작 방향은 팔꿈치로부터 손목 관절의 방향일 수 있다. 깊이 이미지의 근접 화소들은, 팔꿈치로부터 손가락 끝으로 이동하는 포인트들에 대해 우선권(preference)을 부여하기 위한 방법으로써 시작 방향의 투사(projection)에 기반하여, 반복적으로 스코어가 매겨질 수 있고, 한편, 깊이 불연속점과 같은 깊이 계속성(consistency) 제약을 사용하여 깊이 이미지에서 사용자의 손의 경계 또는 극값(extreme value)이 식별될 수 있다. 몇몇 실시예에서, 시작 방향의 양의 방향과 음의 방향 모두에서 깊이 지도 탐색(depth map search)을 제한하기 위해, 고정 값에 기반하거나 또는 사용자의 추정 크기에 기반하여 크기가 변경되는 임계 거리 값이 사용될 수 있다.
- [0043] 또 다른 예를 들면, 깊이 불연속점까지 깊이 이미지의 모든 화소를 포함시키기 위해, 골격 관절(예컨대, 손목 또는 손가락 끝 관절)에 기반하여 배치되는 바운딩 구(bounding sphere) 또는 적절한 다른 바운딩 형상을 사용할 수 있다. 예를 들어, 깊이 이미지의 손 영역의 경계를 짓는 데 사용될 수 있는 깊이 불연속점을 식별하기 위해, 바운딩 구 위로 창(window)을 슬라이딩 할 수 있다.
- [0044] 반복적으로 식별될 수 있는 손 바닥의 중심 주변에 바운딩 형상을 배치하기 위해 바운딩 형상 방법을 사용할 수도 있다. 이런 반복적인 바운딩 방법의 일례가 첨부 문서 1로 본원에 첨부된 "Kinect Developer Summit at GDC 2011: Kinect for XBOX 360"이라는 제목의 David Tuft의 프리젠테이션 및, IEEE International Conference on Systems, Man, and Cybernetics, vol. 2, 2000에서의 "3D drawing system via hand motion recognition from cameras"라는 제목의 K. Abe, H. Saito, S. Ozawa의 논문에 개시되어 있으며, 이 논문은 그 전체가 참조로서

본원에 통합된다.

[0045] 일반적으로, 이런 방법은 모델에서 화소들을 추려내기(cull) 위해 여러 반복 단계를 포함한다. 각각의 단계에서, 상기 방법은 손 중심에 있는 구 또는 다른 형상의 바깥에 있는 화소들을 추려낸다. 다음으로, 상기 방법은 손의 끝으로부터 너무 멀리 있는 화소들을 (팔 벡터에 따라) 추려낸다. 다음으로 상기 방법은 에지 검출 단계(edge detection step)를 실행하여, 손 경계를 에지 검출하고 연결되지 않은 섬(unconnected islands)을 제거한다. 이런 방법의 예시적인 단계들이 도 8의 순서도에 도시되어 있다. 단계(224)에서, 골격 인식 엔진(192)으로부터의 손 관절 데이터에 의해 손 중심 주변에 바운딩 형상이 생성된다. 바운딩 형상은 전체 손을 둘러쌀 정도로 충분히 크며 삼자원이다. 단계(226)에서, 바운딩 형상 바깥에 있는 화소들이 추려내진다.

[0046] 깊이 이미지에서, 사용자의 손이 사용자의 신체 또는 사용자의 다른 손에 가까이에 있어서, 그 다른 신체 부위로부터의 데이터가 세그먼트화된 이미지에 처음부터 포함될 경우가 있을 수 있다. 연결된 컴포넌트 레이블링(labeling)이 실행되어, 세그먼트화된 이미지의 여러 다른 중심들을 레이블링할 수 있다. 중심이, 주로 손이 그 크기와 손 관절의 위치에 기반하여 선택된다. 선택되지 않은 중심들은 추려내질 수 있다. 단계(230)에서, 연결된 팔로부터의 벡터를 따라 손의 끝에서 너무 멀리 떨어져 있는 화소들도 추려내질 수 있다.

[0047] 골격 인식 엔진(192)으로부터의 골격 데이터에는 노이즈가 있을 수 있어서, 손의 중심을 식별하기 위해 손에 대한 데이터를 추가적으로 정제한다. 이는, 이미지를 반복 처리하고 각각의 화소로부터 손의 실루엣(silhouette)의 에지까지의 거리를 측정함으로써 실행될 수 있다. 다음으로, 이미지 세그멘테이션 엔진(194)은 각종 평균을 내어 최대/최소 거리를 계산할 수 있다. 즉, 단계(232)에서, 세그먼트화된 손 이미지에서의 각각의 화소에 대해, x 축과 y 축을 따라 손 실루엣의 에지까지의 최대 거리와 x 축과 y 축을 따라 손 실루엣의 에지까지의 최소 거리가 식별된다. 에지까지의 거리를 각종치로 간주하고, 측정된 모든 화소에 대해 최소로 결정된 거리의 각종 평균을 내어 이미지 내에서 손 위치의 그럴법한 중심(lucky center)을 계산한다(단계(234)). 새로운 중심을 사용하여, 이전 반복으로부터의 손바닥 중심의 변화가 허용 오차 범위 안에 있을 때까지 상기 과정이 되풀이하여 반복될 수 있다.

[0048] 다른 접근 방법으로, 사용자가 손을 상체의 바깥으로 또는 위로 또는 앞으로 들어올릴 때 손 영역의 세그먼트화가 실행될 수 있다. 이렇게 하여, 손 영역이 신체와 더 쉽게 구별될 수 있기 때문에 깊이 이미지에서 손 영역의 식별이 덜 애매해질 수 있다. 사용자 손의 손바닥이 캡쳐 장치(20)를 향하여 그 손의 피처(feature)가 실루엣으로써 검출될 수 있을 때 손 이미지가 특히 설명해진다. 피처에는 노이즈가 있을 수 있지만, 실루엣이 나타난 손을 통해, 예컨대, 손가락 사이의 간격을 검출하고, 손의 전반적인 형상을 확인하고, 다양한 여러 다른 접근 방법을 사용하여 이를 매핑하는 것에 기반하여, 손이 무엇을 하고 있는지에 대한 숙고된 결정(informed decision)을 할 수 있다. 이를 간격 및 다른 피처들을 검출함으로써 특정 손가락의 인식 및 그 손가락이 가리키는 곳의 일반적인 방향을 인식할 수 있다.

[0049] 전술한 손 세그멘테이션 예시들은 예를 들기 위한 것으로, 본 개시의 범위를 제한하기 위한 것은 아님을 이해할 것이다. 일반적으로, 임의의 손 또는 신체 부위 세그멘테이션 방법이 단독으로 또는 서로 결합하여 및/또는 전술한 예시적인 방법들 중 하나와 결합하여 사용될 수 있다.

[0050] 도 6의 파이프라인을 계속 참조하면, 단계(210)는 영역에 대한, 예컨대, 단계(208)에서 식별된 손에 대응하는 깊이 이미지의 영역에 대한 형상 서술자(shape descriptor)를 추출하는 단계를 포함한다. 단계(210)에서의 형상 서술자는 서술자 추출 엔진(196)에 의해 추출되며, 이는 손 영역을 구분하기 위해 사용되는 손 영역의 임의의 적절한 표현일 수 있다. 몇몇 실시예에서, 형상 서술자는 손 영역의 형상을 부호화하거나(codify) 설명하는 벡터 또는 일련의 숫자일 수 있다.

[0051] 서술자 추출 엔진(196)은 단계(210)에서 임의의 다양한 필터를 사용하여 형상 서술자를 추출할 수 있다. 하나의 필터가 화소 분류기(pixel classifier)라고 불리기도 하며, 이는 도 9의 순서도, 도 10의 결정 트리(decision tree)와 도 11-14의 도시를 참조하여 이제부터 설명될 것이다. 단계(240)에서, 세그먼트화된 이미지의 전경에 있는 화소가 선택된다. 이를 화소는 적어도 명목상으로 사용자의 손의 일부라고 생각되는 화소들이다. 중심에 선택된 화소와 함께, 선택된 화소 주변에 사전 정의된 크기의 상자가 취해진다. 실시예에서, 상자의 크기는 정규화된(normalized) 손가락의 너비(width)의 1.5배로 선택될 수 있다. "정규화된 손가락"이란 골격 모델의 크기 및 캡쳐 장치(20)와 사용자의 검출된 거리에 기반하여 정규화된 크기로 조정된 사용자의 손가락이다. 명목상 손의 일부라고 여겨지는 각각의 화소에 대해 다음의 단계들이 연속적으로 실행된다.

[0052] 단계(242)에서, 화소 분류기 필터는 상자의 얼마나 많은 에지가 교차하는지를 결정한다. 교차점은 이미지가 (손

위의) 전경에서 (손 위가 아닌) 배경으로 전환되는 곳이다. 예를 들어, 도 11a는 손가락(276), 손가락에서 선택된 화소(278), 및 반경 r로 화소 주변에 있는 전술한 상자(280)를 도시한다. 상자는 하나의 에지를 따라 있는 두 개의 포인트, 포인트(281a 및 281b)에서 교차한다. 포인트(281a, 281b)는 이미지가 전경(손)에서 배경으로 전환하는 곳이다. 이하에서 설명되는 바와 같이, 손 중심을 정의하기 위해서, 각각의 상자(280)의 에지와 두 개의 교차점을 갖고 있는 모든 화소들(278)은 손가락 끝(또는 아래에서 설명되는 바와 같이 손가락 관절(knuckle) 또는 팔의 일부)으로 간주된다.

[0053] 단계(246)에서, 화소 분류기 필터는 교차점들이 동일한 에지 또는 다른 에지에 있는지를 결정한다. 도 11b에 도시된 바와 같이, 손가락은 동일한 에지가 아닌 인접한 두 에지를 따라 상자(280)와 교차할 수 있다. 아래에 설명되어 있는 대로, 이 정보는 손가락이 가리키는 방향을 결정하는 데에 사용될 것이다.

[0054] 손가락 끝과는 대조적으로, 아래에서 설명되는 바와 같이, 상자(280)와 네 점에서 교차하는 화소는 손 중심을 정의하기 위한 손가락으로 간주될 것이다. 예를 들어, 도 12는 선택된 화소(278)가 손가락 끝과 충분히 멀리 떨어져 있고, 상자(280)와 네 개의 교차점(281a, 281b, 281c 및 281d)이 있는 일례를 도시한다.

[0055] 도 9의 순서도의 단계(242)와 도 10의 결정 트리의 264에서, 화소 분류기 필터는 상자(280)의 몇 개의 에지에서 교차되고 있는지를 확인한다. 아무런 에지에서도 교차되지 않으면, 선택된 화소는 사용자의 손바닥 안에 있는 것으로 265에서 간주된다. 즉, 화소가 손가락 안에 또는 손가락 끝에 있으면 적어도 두 개의 에지에서 교차하고 화소가 손에 있으면 아무런 에지에서도 교차하지 않도록 상자(280)의 크기가 선택되기 때문에, 화소는 손바닥에 있는 것으로 여겨진다. 두 개의 에지에서 교차되면, 이하에서 설명되는 대로 필터는 266으로 진행하여 교차점이 없는(non-intersected) 에지의 모서리(corner)들이 채워져 있는지(solid)(손 위) 또는 비어 있는지(배경)를 확인한다. 267에서 네 개의 에지에서 교차되면, 전술한 바와 같이 이는 손가락으로 간주된다. 268에서 상자(280)의 에지들이 여섯 번 교차하면, 이는 유효하지 않은 판독으로 간주되어 폐기된다(단계(250)).

[0056] 두 개의 에지에서 교차되는 다시 266을 참조하면, 이는 손가락 끝일 수도 있지만, 두 인접한 손가락들 사이의 공간일 수도 있다. 화소 분류기 필터에서 교차점이 없는 에지의 모서리들을 확인한다(단계(248)). 교차점이 없는 에지의 모서리들이 채워져 있으면, 이는 상자가 이를 모서리에서 손 위에 위치하고 교차점은 인접 손가락들 간의 골(valley)을 나타내는 것을 의미한다. 역으로, 교차점이 없는 에지의 모서리들이 비어 있으면(266과 관련된 도시에서 보여짐), 이는 상자가 이를 모서리에서 배경에 위치하고 교차점은 손의 일부를 나타내는 것을 의미한다.

[0057] 모서리들이 비어 있으면, 화소 분류기 필터는 교차점들 간의, 코드(chord) 길이라고도 하는 거리가 손가락의 최대 너비보다 작은지를 269에서 확인한다(단계(252)). 즉, 두 개의 교차점이 있는 경우, 이는 도 11a에 도시된 바와 같이 손가락 끝일 수도 있다. 반면, 화소는, 도 13에 도시된 바와 같이, 팔의 일부, 또는 손가락 관절과 같은 손의 일부일 수도 있다. 이런 경우, 코드의 길이(282)는 손가락의 최대 너비보다 클 수 있다. 이런 경우, 상자(280)가 검사 중인 화소(278)는 271에서 팔 또는 손가락 관절에 위치한다고 본다(도 10).

[0058] 손가락 끝 또는 손가락을 식별하는 것은 물론, 두 개의 포인트 교차 또는 네 개의 포인트 교차는 손가락 끝/손가락이 가리키고 있는 방향도 나타낼 수 있다. 예를 들어, 도 11a에서, 손가락의 최대 너비보다 작은 두 개의 교차점이 있어서, 화소(278)는 손가락 끝에 위치한다고 결정되었다. 하지만, 이런 교차점이 주어지면, 손가락 끝이 가리키고 있는 방향에 대해 추론할 수 있다. 도 12에 도시된 손가락에 대해서도 마찬가지일 수 있다. 도 11a는 위를 똑바로 가리키는 손가락(276)을 도시한다. 그러나, 손가락 끝(276)은 다른 윗 방향을 가리키고 있을 수도 있다. 손가락 끝(276)에서 포인트(278) 근처에 있는 다른 포인트들로부터의 정보를 사용하여 방향에 대해 더 추론할 수도 있다.

[0059] 도 11b는 손가락/손가락 끝이 가리키고 있는 방향에 대한 추가적인 추론을 제공하는 두 개의 포인트 교차를 도시한다. 즉, 공유 모서리까지의 거리의 비율로부터 방향을 추론할 수 있다. 다른 식으로 말하면, 포인트들(281a 및 281b) 간의 코드 길이는 포인트들(281a 및 281b)과 공유 모서리 간의 변들을 포함하여 삼각형의 빗변을 정의한다. 손가락이 빗변에 수직인 방향을 가리키고 있다고 추론할 수 있다.

[0060] 손이 두 개의 손가락을 함께, 세 개의 손가락을 함께, 또는 네 개의 손가락을 함께 붙이고 있는 경우가 있을 수 있다. 따라서, 손에 있는 각각의 화소에 대해 상자(280)를 사용하여 상기의 단계들을 실행한 후에, 두 손가락을 합한 최대 너비보다 약간 더 큰 상자(280)를 사용하여 상기 과정을 반복하고, 다시 세 손가락을 합한 최대 너비보다 약간 더 큰 상자(280)를 사용하여 상기 단계를 반복할 수 있다.

[0061] 화소 분류기 필터 데이터가 모아지면, 다음으로 화소 분류기 필터는 단계(258)에서 데이터로부터 손 모델을 구

축하려고 한다(도 9). 이런 정보로부터, 아래에서 설명되는 바와 같이 분류기 엔진(198)은 손이 펼쳐져 있는지 또는 오므려져 있는지를 결정할 수 있다. 예컨대, 손가락 끝인 영역, 및 손바닥인 영역, 손 세그멘테이션 단계로부터의 손바닥 중심으로 추측되는 영역 등 식별된 작은 영역들, 또는 중심이 있다. 다음으로 분류기 엔진(198)은, 네 개의 포인트에서 교차했다는 사실 때문에 손가락 끝으로 분류되지 못하고 손가락으로 분류된 손가락 중심들을 검사한다. 손가락 및 손가락 끝 영역에 대한 방향성(directional orientation)도 식별되었다. 손가락 중심이 손가락 끝 중심과 나란히 있고 이들이 서로에 대해 적절한 상대적인 위치에 있으면, 알고리즘은 동일한 손가락에 속한다고 하여 이 중심들을 연결한다.

[0062] 다음으로, 골격 크기, 및 손가락이 어느 정도의 크기라고 여겨지는 것에 기반하여, 손가락의 관절이 있다고 여겨지는 곳이 손가락 영역의 방향(orientation)을 사용하여 투사된다. 결정된 손 모델을 확정하기 위해 임의의 식별된 손가락들 사이의 골의 크기, 위치 및 방향이 사용될 수도 있다. 다음으로, 투사된 손가락 관절의 위치가 손바닥에 연결된다. 이를 마치면, 화소 분류기 엔진은 골격 손 모델(284)을 결정하고, 그 중 두 개의 예시가 도 14에 도시된다. 모델은, 손바닥 중심에 연결된, 손가락 관절 중심에 연결된, 손가락 중심에 연결된 손가락 끝 중심을 포함한다. 공지된 기하학 및 공지된 팔의 위치로부터 가능한 손의 위치에 관한 데이터를 사용하여 손가락 끝, 손가락, 손가락 관절 및/또는 손바닥 중심 위치의 결정된 위치들을 입증하거나 반박할 수 있을 뿐만 아니라 손의 일부를 형성하지 않는 것으로 결정될 수 있는 중심 데이터를 폐기할 수도 있다.

[0063] 손의 하나 이상의 부분이 모델에서 누락되어도 상기로부터 손 모델을 구성할 수 있다. 예를 들어, 손가락이 가려져 있거나, 사용자의 신체나 다른 손에 너무 가까워서 검출하지 못할 수 있다. 또는 사용자가 손가락이 없을 수도 있다. 화소 분류 필터는, 필터에서 검출하는 손가락과 손의 위치를 사용하여 손 모델을 구성할 수 있다.

[0064] 화소 분류 필터와 함께 또는 필터를 대신해서 실행될 수 있는 다른 필터로 곡률 분석(curvature analysis) 필터를 들 수 있다. 이 필터는 세그먼트화된 손 실루엣의 경계를 따르는 곡률에 중점을 두어 손가락을 구별하기 위해 정점(peak)과 골(valley)을 결정한다. 도 15의 순서도를 참조하면, 단계(286)에서, 첫 번째 화소에서 시작하여, 여덟 개의 주변 화소들을 검사하여 손에서 어떤 화소가 옆의 화소인지를 결정한다. 그 화소와 그 옆의 화소 간의 연결을 위해서 각각의 화소에 0에서 7 사이의 값이 부여된다. 손 경계를 제공하는 손의 실루엣 주변에 이들 숫자들의 체인(chain)이 만들어진다. 이 값들은 단계(288)에서 각도와 윤곽선(contour)으로 전환되어, 도 16에 도시된 바와 같이, 손 윤곽선과 정점의 그래프를 제공할 수 있다. 손 윤곽선과 정점을 생성하는 이와 같은 단계들은 예컨대, Computer Vision and Robotics Laboratory, McGill University, Montreal, Quebec, Canada, 1988 안의 "Curvature morphology"라는 제목의 F. Leymarie, M.D. Levine의 논문에 설명되어 있고, 이 논문은 그 전체가 참조로서 본원에 통합된다.

[0065] 단계(289)에서 손 실루엣 주변의 정점이 식별되고, 정점의 다양한 피처(feature)에 대해서 각각의 정점이 분석된다. 정점은 시작점, 정점과 종점으로 정의될 수 있다. 아래에서 설명되는 바와 같이 이들 세 포인트는 삼각형을 형성할 수 있다. 검사 가능한 정점의 다양한 피처(feature)는 예를 들어,

[0066] · 정점의 너비;

[0067] · 특정 정점의 최대 높이;

[0068] · 정점 내에서 곡률 샘플들의 평균 높이;

[0069] · 정점 형상 비율(최대 높이/평균 높이);

[0070] · 정점의 면적;

[0071] · 손에서 정점까지의 거리;

[0072] · 팔꿈치에서 손까지의 방향(x, y 및 z);

[0073] · 정점 방향과 팔 방향의 외적(cross product)(팔 방향과 정점 방향 사이의 각도가 얼마나 작은지 여부); 및

[0074] · 정점 시작점과 최대점(max point) 간의 벡터 및 최대점과 종점 간의 벡터의 외적을 포함한다.

[0075] 이 정보는, 손가락과 손을 구별하기 위해, 예컨대, 서포트 벡터 머신(support vector machine)과 같은 다양한 기계 학습 기법(machine learning technique)을 통해 단계(290)에서 사용될 수 있다. 서포트 벡터 머신은 예를 들어, C. Cortes and V. Vapnik의 논문 Support-Vector Networks, Machine Learning, 20(3):273-297, September 1995, 및 Vladimir N. Vapnik의 논문 The Nature of Statistical Learning Theory. Springer, New York, 1995에서 알려져 설명되어 있으며, 이들은 그 전체가 참조로서 본원에 통합된다. 다른 기계 학습 기법이

사용될 수 있고, 예를 들어, 회귀(regression), 최근린(nearest neighbor), (지도되지 않은)지도된 클러스터링((un)supervised clustering) 등을 포함할 수 있다. 실시예에서, 히든 마르코프 모델(Hidden Markov Model)을 사용하여 노이즈가 있는 데이터를 스무딩하여(smoothed), 손의 상태를 유지하고 노이즈를 필터링할 수 있다. 꼭 를 분석 필터는 캡쳐 장치(20)를 바라보는 펼친 손을 식별하는 데 잘 동작하지만, 오므린 손을 식별하는 데에도 사용될 수 있다.

[0076] 전술한 필터들은 손의 실루엣에 관한 데이터를 검사한다는 점에서 실루엣 필터라고 부를 수 있다. 사용 가능한 다른 필터로는 히스토그램(histogram) 필터가 있고, 손 모델을 구성하기 위해 깊이 데이터를 사용한다는 점에서 깊이 필터라고도 한다. 이 필터는 전술한 필터들에 더해서 또는 필터들을 대신해서 사용될 수 있고, 사용자가 자신의 손을 이미지 캡쳐 장치(20) 쪽으로 가리키고 있을 때 특히 유용할 수 있다.

[0077] 히스토그램 필터에서, 손 영역에서의 거리의 히스토그램을 구성할 수 있다. 예를 들어, 이런 히스토그램은 열다섯 개의 빈(bin)을 포함할 수 있고, 여기서 각각의 빈은, 카메라에 가장 가까운 포인트로부터 Z-방향(깊이)으로의 거리가 그 빈과 관련된 특정 거리 범위 내에 있는 손 영역의 포인트들의 개수를 포함한다. 예를 들어, 이런 히스토그램에서의 첫 번째 빈은 손 중심까지의 거리가 0에서 0.40 센티미터 사이에 있는 손 영역의 포인트의 개수를 포함하고, 두 번째 빈은 손 중심까지의 거리가 0.40에서 0.80 센티미터 사이에 있는 손 영역의 포인트의 개수를 포함할 수 있다. 이런 방식으로, 손의 형상을 부호화할 벡터가 구성될 수 있다. 또한 이런 벡터는 예를 들어, 추정된 신체 사이즈에 기반하여 정규화될 수 있다.

[0078] 일반적으로, 빈에서의 분포를 분석함으로써, 캡쳐 장치(20)를 가리키는 손이 펼쳐져 있는지 오므려져 있는지를 결정할 수 있다. 분포가 앞쪽에 많으면(처음 몇 개의 빈이 마지막 몇 개의 빈보다 많은 수의 화소를 갖고 있음), 손을 오므리고 있다고 추론할 수 있다. 분포가 뒤쪽에 많으면(마지막 몇 개의 빈이 처음 몇 개의 빈보다 많은 수의 화소를 갖고 있음), 손이 펼쳐져 있고 캡쳐 장치(20)를 가리키고 있다고 추론할 수 있다. 따라서, 히스토그램이 나타내는 질량 분포(mass distribution)로부터 펼친 또는 오므린 상태가 추론될 수 있다. 또는, 히스토그램은, 손 상태를 자동으로 예측하기 위한 학습에 관한 다양한 기계 학습 분류 기법(본원에서 설명됨)에서 피처(feature)로써 사용될 수 있다.

[0079] 예시적인 다른 접근 방법으로, 손 영역의 포인트로부터, 사용자의 추정된 골격의 관절, 뼈 세그먼트 또는 손바닥 평면, 예컨대, 팔꿈치 관절, 손목 관절 등까지의 거리 및/또는 각도에 기반하여 히스토그램을 구성할 수 있다. 도 17은 오므린 손과 펼친 손에 대해 결정된 히스토그램을 나타내는 두 그래프를 도시한다.

[0080] 또 다른 필터는 푸리에 서술자(Fourier descriptor)이다. 푸리에 서술자의 구성은, 예를 들어, 손 영역의 각각의 화소로부터 손 영역의 주변까지의 거리를 손의 경계의 타원형 맞춤(eliptical fitting)의 반경에 매핑함으로써 손 영역의 윤곽선을 부호화한 후, 지도에 푸리에 변환을 실행하는 단계를 포함할 수 있다. 또한, 이런 서술자들은 예컨대, 추정된 신체 사이즈에 따라 정규화될 수 있다. 이런 서술자들은 이동(translation), 크기 조정(scale), 및 회전에 따라 변하지 않을 수 있다.

[0081] 또 다른 필터로는 손 영역의 볼록 헬(convex hull)에 대한 손 영역의 윤곽선 영역의 비율을 결정함으로써 손 영역의 볼록성(convexity)을 결정하는 필터를 포함한다.

[0082] 상기의 필터들이 불충분한 정보를 제공하는 경우에 유용할 수 있는 추가적인 필터를 불연속점 필터(discontinuity filter)라고 한다. 이런 필터는 인식 또는 위치 정보를 제공하지 않고, 이미지를 손이라고 가정하여 손을 펼치고 있는지 오므리고 있는지를 식별한다. 불연속점 필터는, 오므린 주먹은 펼친 손의 윤곽선 보다 상대적으로 부드럽고 적은 윤곽선 불연속점을 갖고 있다는 점에 기반한다. 따라서, 측정된 윤곽선 불연속점이 사전 결정된 임계치보다 적으면, 불연속점 필터는 오므린 손이라고 식별한다. 측정된 윤곽선 불연속점이 사전 결정된 임계치보다 많으면, 불연속점 필터는 펼친 손이라고 식별한다. 이런 이미지 데이터의 예시가 도 18에 도시된다.

[0083] 손의 펼친/오므린 상태를 검출하는 추가적인 필터를 내추럴 그랩 필터(natural grab filter)라 한다. 일반적으로, 내추럴 그랩 필터는 사용자가 캡쳐 장치(20)로 손을 뻗어 자신의 손을 펼치고 오므리는 움직임을 검출한다. 손바닥이 캡쳐 장치(20)를 마주보거나 마주보지 않을 수도 있다. 내추럴 그랩 필터의 추가적인 설명은 도 19의 순서도를 참조하여 설명된다. 단계(291)에서, 추가적인 세그멘테이션이 실행된다. 이런 추가적인 세그멘테이션은 아래팔을 설명하는 골격 인식 엔진(192)에서 제공되는 엘보우-투-핸드 벡터(elbow-to-hand)를 검사한다. 내추럴 그랩 필터는 현재 손 센그멘테이션 상에서 각각의 화소에 대해 반복되어, 화소가 팔꿈치로부터 얼마나 멀리 있는지를 알아낸다. 팔꿈치로부터 가장 멀리 있는 화소들만이 유지되고, 여기서 가장 멀리 있는 화소는 사

용자의 특정 골격 크기에 기반하는 창(window)으로써 정의된다. 예를 들어, 성인에 대해서는 팔꿈치로부터 가장 멀리 있는 포인트에서 시작하는 4 인치 창일 수 있다. 이 창은 어린이 또는 더 작은 사람의 경우 변경될 수 있다.

[0084] 이 단계에서 세 개의 서브-필터 중 하나 이상이 실행된다. 첫 번째 필터는 깊이 변화, 또는 점프를 측정한다. 단계(292)에서, 깊이 점프 서브-필터는 최근에 세그먼트화된 손 이미지의 각각의 화소 행(row) 및 열(column)에서 반복된다. 이런 반복은 모든 행 다음에 모든 열, 또는 그 반대로 실행될 수 있다. 이 서브-필터는 깊이 점프의 개수를 세고, 여기서 깊이 점프는 하나의 화소에서 다음 화소로 이미지에서 나타나는 깊이에서의 큰 점프라고 여겨진다. 모든 행 및 열에 대해 반복된 후에, 서브-필터는 깊이 점프가 있는 행 및 열의 백분율을 찾아낸다.

[0085] 이 백분율에 기반하여, 깊이 점프 서브-필터는 손이 펼쳐져 있는지 오므려져 있는지의 타당한 추정을 할 수 있다. 펼친 손은 손가락 사이의 공간에서 깊이 점프가 있을 것이기 때문에, 오므린 손은 펼친 손보다 상당히 적은 수의 깊이 점프를 갖고 있을 것이다(깊이 점프는, 예를 들어, 카메라 데이터 노이즈로 인한 펼친 손가락과 같이 다른 이유들로 인해 생길 수도 있다). 깊이 점프 서브-필터의 출력은 백분율이므로, 손 크기에 따라 변하지 않는다. 출력은 테스트에서 생성된 사전 정의된 임계 백분율에 기반하는 이진값(binary)(펼침/오므림)일 수 있다. 이 대신에 또는 추가적으로, 출력은 신뢰도 값(confidence value)(백분율이 임계치에 얼마나 가까운지 여부)을 갖는 펼친/오므린 출력일 수 있다.

[0086] 두 번째 서브-필터는 원형(circle) 서브-필터라고 할 수 있다. 단계(293)에서, 원형 서브-필터는 실루엣이 전반적으로 원형인지 여부를 결정하기 위해 최근에 세그먼트화된 손 이미지의 실루엣을 검사할 수 있다. 전반적으로 원형인지, 이는 오므린 손을 나타낸다. 세그먼트화된 이미지가 손가락 또는 카메라의 추가 노이즈를 포함하든 해서 손이 펼쳐진 경우, 덜 둥글 것이다.

[0087] 이런 결정을 내리는 방법으로 최근에 세그먼트화된 손의 질량 중심을 계산한 후 그 이미지의 실루엣의 반경의 표준 편차를 조사한다. 오므린 주먹은 원형에 가깝고 중심과 실루엣의 가장자리까지의 다른 반경들의 표준 편차가 상대적으로 적을 것이다. 표준 편차는, 비원형인 것에 더 많은 변화가 있기 때문에 더 클 것이다. 따라서, 표준 편차 값에 기반하여, 원형 서브-필터는 손이 펼쳐져 있는지 오므려져 있는지의 추정치를 출력한다. 이 필터는 또한, 손이 캡처 장치로부터의 광선에 수직인 실루엣에 있거나 또는 캡처 장치를 가리키고 있을 때에도 동작한다. 원형 서브-필터의 출력은 테스트에서 생성된 사전 정의된 임계 표준 편차에 기반하는 이진값(펼침/오므림)일 수 있다. 이 대신에 또는 추가적으로, 출력은 신뢰도 값(표준 편차가 임계치에 얼마나 가까운지 여부)을 갖는 펼친/오므린 출력일 수 있다.

[0088] 세 번째 서브-필터는 패들(paddle) 서브-필터라고 할 수 있다. 단계(294)에서, 패들 서브-필터는 사용자가 자신의 손 또는 손가락을 뻗었지만, 따라서 펼친 손이지만, 모든 손가락이 함께 붙어서 기본적으로 패들을 형성하는지를 조사한다. 패들 서브-필터는 손가락 끝에서부터 팔꿈치 쪽으로 빈을 갖고 있는 히스토그램을 생성하고, 빈 각각에서 화소의 개수를 조사한다. 다음으로 빈이 점점 커지고 있는지, 또는 일정하다가 결국 작아지고 있는지를 분석한다.

[0089] 손가락을 함께 붙여서 손을 패들처럼 들고 있을 때, 손가락 끝에는 더 적은 수의 화소들이 있고 손의 중심으로 갈수록 더 많은 화소들이 있을 거라는 생각이다. 역으로, 손이 주먹을 쥐고 있는 경우, 주먹이 있는 맨 위에 상대적으로 더 많은 화소들이 있고 손목과 팔로 갈수록 점점 줄어들 것이다. 따라서, 패들 서브-필터의 히스토그램을 사용하고, 패들 대 주먹일 때의 손의 공지된 구성을 사용하여 획득한 빈을 분석함으로써, 패들 서브-필터는 세그먼트화된 손이 펼쳐져 있는지 또는 오므려져 있는지의 추정치를 제공할 수 있다. 다시, 패들 서브-필터의 출력은 테스트에서 생성된 사전 정의된 임계치에 기반하는 이진값(펼침/오므림)일 수 있다. 이 대신에 또는 추가적으로, 출력은 신뢰도 값(값이 임계치에 얼마나 가까운지 여부)을 갖는 펼친/오므린 출력일 수 있다.

[0090] 단계(295)에서, 서브-필터로부터의 결과가 결합되어 손이 펼쳐져 있는지 또는 오므려져 있는지의 최종 출력 또는 추정치를 얻을 수 있다. 단계(295)의 결합은, 이를 각각의 필터에 대한 펼쳐짐과 오므림 및 신뢰도 값을 보고 최종 펼쳐짐/오므림 출력을 위해 이들을 결합하는 결정 트리를 사용한다. 결정 트리는 각각의 서브-필터의 장점을 극대화하고 약점을 최소화하도록 커스터마이즈될 수 있다. 깊이 점프 서브-필터가 가장 신뢰할 수 있다. 따라서, 깊이 점프 서브-필터의 출력이 임계치보다 높은 신뢰도를 가질 때, 이는 최종 출력으로 취해질 수 있다.

[0091] 원형 서브-필터의 결과가 그 다음으로 신뢰할만하다. 깊이 점프 서브-필터의 결과가 임계치보다 높은 신뢰도를

갖고 있지 않은 경우, 또는 깊이 점프 서브-필터의 결과와 원형 서브-필터의 출력이 서로 맞지 않는 경우, 원형 서브-필터의 결과가 그 신뢰도 값에 따라 고려될 수 있다. 고신뢰도 값이면, 원형 서브-필터의 출력이 최종 출력으로 취해질 수 있다.

[0092] 패들 서브-필터의 결과가 그 다음으로 신뢰할만하다. 깊이 점프 또는 원형 서브-필터 중 어느 필터도 신뢰할만한 결과를 제공하지 않으면, 패들 필터의 결과가 고려되고, 임계치보다 높은 신뢰도 값을 갖고 있으면 최종 출력으로 사용될 수 있다.

[0093] 동일한 결정 트리-유형의 분석을 사용하여 전술한 다른 필터들의 결과를 고려할 수 있다. 일 실시예에서, 곡률 분석 필터가 사전 결정된 임계치보다 높은 신뢰도로 출력을 제공하는 경우, 그 결과가 최종 출력으로 취해질 수 있다. 곡률 분석 필터의 신뢰도가 임계치보다 높지 않으면, 히스토그램 필터의 신뢰도가 조사될 수 있다. 사전 결정된 임계치보다 높으면, 그 결과가 최종 출력으로 취해질 수 있다. 히스토그램 필터의 신뢰도가 임계치보다 높지 않으면, 내추럴 그랩 필터의 신뢰도가 조사될 수 있다. 사전 결정된 임계치보다 높으면, 그 결과가 최종 출력으로서 취해질 수 있다.

[0094] 전술한 필터들은 다른 크기의 손에 따라 정규화될 수 있고 및/또는 변하지 않을 수 있다. 다른 필터들도 다수의 플레이어, 및 이런 각각의 플레이어의 한 손 또는 양손으로 확장될 수 있다.

[0095] 형상 서술자의 필터 예시들은 단지 예를 들기 위한 것에 불과하며 본 개시의 범위를 제한하기 위한 것이 아님을 이해할 것이다. 일반적으로, 손 영역에 대한 임의의 적절한 형상 서술자가 단독으로 또는 서로 결합하여 및/또는 전술한 예시적인 방법들 중 하나와 결합하여 사용될 수 있다. 예를 들어, 전술한 히스토그램 또는 백터들과 같은 형상 서술자는 더 큰 백터들로 섞어 맞추기(mix and match), 결합, 및/또는 이어 붙이기(concatenate)를 할 수 있다. 이로써, 따로 떼어서 보는 것으로는 식별할 수 없었던 새로운 패턴의 식별을 할 수 있다. 이를 필터는, 예컨대, 식별된 손가락이 이전 프레임에서 식별된 손가락에서 너무 많이 벗어나 있는지를 나타내 줄 수 있는 과거 프레임 데이터의 사용함으로써 향상될 수 있다.

[0096] 도 6의 도시된 파이프라인에 이어서, 전술한 필터들로부터의 정보를 사용하여, 분류기 엔진(198)(도 2)은 단계(214)에서 손을 분류할 수 있다. 예를 들어, 단계(210)에서 추출된 형상 서술자는 저장된 트레이닝 데이터(training data)에 기반하여 분류되어 손의 상태를 추정할 수 있다. 예를 들어, 도 7의 310에 도시된 바와 같이, 손은 펼치거나 오므린 것으로 분류될 수 있다. 몇몇 예시에서, 트레이닝 데이터는 다양한 손 상태의 레이블링된(labeled) 깊이 이미지 예시들을 포함할 수 있다. 트레이닝 데이터는 실제하는 것이거나 또는 합성하여 생성될 수 있고, 다른 신체 사이즈와 팔 방향을 갖고 있는 전신 또는 상체 3D 모델들을 묘사하는 것은 물론, 모션 캡처 또는 손으로 디자인한 손 포즈에 기반하여 다르게 연결된 손 포즈들도 묘사한다.

[0097] 몇몇 예시에서, 단계(214)의 분류에서 사용된 트레이닝 데이터는 사전-결정된 손 예시들의 집합에 기반할 수 있다. 손 예시들은, 손 영역에 관한 형상 서술자와 비교되는 대표 손 상태에 기반하여 그룹화되거나 레이블링될 수 있다.

[0098] 몇몇 예시에서, 트레이닝 데이터를 구분하기 위해 다양한 메타-데이터(meta-data)를 사용할 수 있다. 예를 들어, 트레이닝 데이터는 손의 한 쪽 또는 양쪽(예컨대, 왼쪽 또는 오른쪽), 손 방향(orientation)(예컨대, 아래팔 각도 또는 아래팔 방향), 깊이 및/또는 사용자의 신체 사이즈 중 하나 이상에 기반하여 구분될 수 있는 복수의 손 상태 예시들을 포함할 수 있다. 이를 트레이닝 손 예시들을 분리된 서브셋으로 구분함으로써 각각의 구획 안에서 손 형상의 다양성(variability)을 줄일 수 있고, 이는 손 상태의 보다 정확한 전체적인 분류를 가능케 할 수 있다.

[0099] 추가적으로, 몇몇 예시에서, 트레이닝 데이터는 특정 애플리케이션에 특정되어 있을 수 있다. 즉, 트레이닝 데이터는 특정 애플리케이션에서 예상되는 동작, 예를 들어, 게임에서 예상되는 활동 등에 따라 달라질 수 있다. 또한, 몇몇 예시에서, 트레이닝 데이터는 사용자에 특화되어 있을 수 있다. 예를 들어, 애플리케이션 또는 게임은, 사용자가 하나 이상의 트레이닝 훈련(exercise)을 실행하여 트레이닝 데이터를 보정하는 트레이닝 모듈을 포함할 수 있다. 예를 들어, 사용자는 일련의 퍼고 오므린 손 자세를 실행하여, 시스템과의 차후의 인터랙션 중에 사용자 손의 상태를 추정하는 데 사용되는 트레이닝 데이터 셋을 만들 수 있다.

[0100] 사용자의 손의 분류는 트레이닝 예시에 기반하여 다양한 방식으로 실행될 수 있다. 예를 들어, 분류 시에 다양한 기계 학습 기법을 이용할 수 있다. 비-제한적인 예시로 서포트 벡터 머신 트레이닝, 회귀, 최근린, (지도되지 않은)지도된 클러스터링 등을 포함한다.

[0101] 도 20은 전술한 다양한 필터의 결과를 결합하는 관리자 필터를 도시한다. 예를 들어, 곡률 분석 필터, 깊이 히

스토그램 필터, 도 20에 도시되지 않은 다른 펠친/오므린 손 필터 및 팔 방향에 관련된 데이터가 예컨대, 서포트 벡터 머신과 같은 기계에 제공될 수 있다. 이 결과는 하든 마르코프 모델을 사용하여 추가적으로 정제되어 펠친 또는 오므린 손의 표시를 제공할 수 있다. 앞서 언급한 바와 같이, 사용자의 카메라로부터의 거리 및 분석된 골격에서 추론된 손-크기를 아는 것에 의해, 본원에서 설명된 다양한 필터에서 사용된 실루엣이 손-크기로 크기가 조정되고 센서-거리에는 불변할 수 있다.

[0102] 앞서 논의한 바와 같이, 분류 기법은 관찰된 손의 가능성(likelihood)를 여러 상태의 하나로써 예측하도록 다양한 손 상태의 레이블링된 깊이 이미지 예시들을 사용할 수 있다. 또한, 분류 단계 도중에 또는 분류 단계 이후에 신뢰도가 분류에 추가될 수 있다. 예를 들어, 트레이닝 데이터에 기반하여 또는 분류 단계의 출력에 시그모이드 함수(sigmoid function) 또는 기타 적절한 오차 함수를 피팅하여(fitting), 추정된 손 상태에 신뢰 구간(confidence interval)을 할당할 수 있다.

[0103] 손 상태를 분류하는 간단하고 비-제한적인 예시로써, 도 7의 310에 도시된 바와 같이, 펠치거나 또는 오므리거나 하는, 가능한 두 가지 손 상태가 있을 수 있다. 이와 같이 예시에서, 트레이닝 데이터는 두 개의 레이블링된 손 집합, 즉, 펠쳐져 있거나 또는 거의 펠쳐져 있는 손 상태를 나타내는 손 예시들의 첫 번째 집합과 오므려져 있거나 또는 거의 오므려져 있는 손 상태를 나타내는 손 예시들의 두 번째 집합을 포함할 수 있다. 이렇게 하여, 식별된 손 영역에 대해 추출된 손 서술자가 주어질 때, 식별된 손 영역의 추출된 손 서술자는 첫 번째(펠친) 및 두 번째(오므린) 손 집합의 손 예시들과 비교되어, 식별된 손 영역이 각각의 집합에 속할 가능성을 결정할 수 있다. 그 다음에, 손의 상태는 더 높은 가능성에 기반하여 추정될 수 있다.

[0104] 펠친 또는 오므린 손 상태뿐만 아니라, 예를 들어, 하나 이상의 손가락으로 특정 방향을 가리키는 것과 같이 구체적인 손가락 방향을 식별하는 데에도 본 기술이 사용될 수 있다. 본 기술은 또한, 손가락들이 펠치고 오므린 상태 사이 어딘가에 있는 손 위치, 및 x, y, z 직교 좌표계에서 다양한 각도로 향해 있는 다양한 손의 위치를 식별하는 데에도 사용될 수 있다.

[0105] 예를 들어, 도 7의 310에 도시된 바와 같이, 식별된 손 영역이 펠쳐져 있을 가능성이 더 높다고 결정되어, 따라서 펠쳐져 있다고 분류된다. 또한, 몇몇 예시에서, 식별된 손의 특정 손 상태로 결정된 가능성은 손 상태의 추정에 대한 신뢰도 구간을 생성하는 데 사용될 수 있다.

[0106] 실시예에서, 다양한 분류-후 필터링 단계들을 이용하여 단계(216)에서 손 상태의 추정치의 정확도를 증가시킬 수 있다(도 6). 예를 들어, 시간적 일관성 필터링, 예컨대, 저주파 통과 필터 단계가 연속하는 깊이 이미지 프레임들 사이에서 예측된 손의 상태에 적용되어 예측을 스무딩하고, 예컨대, 가짜 손 움직임, 센서 노이즈, 또는 이따금씩의 분류 에러로 인한 시간적인 지터링(jittering)을 줄일 수 있다. 즉, 캡처 장치 또는 센서로부터의 복수의 깊이 이미지에 기반하는 사용자의 손의 복수 개의 상태가 추정되고, 손의 상태를 추정하기 위해서 복수의 추정치의 시간적인 필터링이 실행될 수 있다. 또한, 몇몇 예시에서, 일부 애플리케이션은 다른 애플리케이션 보다 거짓 긍정(false positive)(한 방향 또는 다른 방향)에 더욱 민감할 수 있어서, 분류 결과가 한쪽 상태 또는 다른 쪽 상태(예컨대, 펠친 또는 오므린 손)로 치우칠 수 있다.

[0107] 단계(220)에서, 도 6의 파이프라인은 추정된 손의 상태에 기반하여 각각의 프레임에서 응답(response)을 출력할 수 있다. 예를 들어, 컴퓨팅 장치(10)의 콘솔(12) 등의 컴퓨팅 시스템의 콘솔로 명령이 출력될 수 있다. 다른 예를 들면, 디스플레이 장치(16) 등의 디스플레이 장치로 응답이 출력될 수 있다. 이렇게 하여, 추정된 사용자의 손 상태를 포함하는 사용자의 추정된 모션이 시스템(10)의 콘솔(12)에 대한 명령으로 번역되어, 전술한 바와 같이 사용자가 시스템과 인터랙션할 수 있다. 또한, 전술한 방법과 프로세스는 예컨대, 입, 눈 등의 사용자의 임의의 신체 부위의 상태의 추정치를 결정하도록 구현될 수 있다. 예를 들어, 전술한 방법을 사용하여 사용자의 신체 부위의 자세를 추정할 수 있다.

[0108] 본 기술을 통해 예컨대, 도 1a 내지 도 1c 도시된 것과 같은 NUI 시스템과 매우 다양한 인터랙션이 가능해진다. 손/손가락 움직임에 기반하거나, 또는 새롭게 인식되는 제스처, 보다 몰입하게 되는 경험 및 흥미진진한 게임 만들기에 바람직한 큰 신체 모션 및 미세한 손 제어를 결합하는, 광범위한 내추럴 인터랙션이 있다. 이를 사용과 인터랙션은 다음을 포함하지만 이에 제한되지는 않는다.

[0109] · 잡기/놓기 기법(Grab/release mechanics) - 사용자의 손이 펠쳐져 있는지 오므려져 있는지를 각각의 프레임에서 정확하게 결정할 수 있음으로써, NUI 시스템 또는 애플리케이션은 잡거나 놓는 모션을 실행할 때 정확하게 결정할 수 있다. 따라서, 예를 들어, 사용자가 사물을 던지는 게임에서, NUI 시스템 또는 애플리케이션은 사용자가 던져지는 사물을 놓으려고 할 때(오므린 손에서 펠친 손으로 전환) 정확한 놓는 시점을 결정할 수 있다.

다른 예시들로, 등산, 슬링샷(sling shot), 풍선 던지기(balloon toss), 및 가상 레버 당기기를 포함하고, 그러나 이에 제한되지는 않는다.

[0110] · 사물 당기기 - 펼친 또는 오므린 손 검출을 통해 가상 사물 - 로프, 레버 등을 당기기.

[0111] · 시작/종료 동작 - 동작을 시작하고 계속하기 위한 펼친 손, 동작을 종료하기 위한 오므린 손. 일례로 게임에서 소방 호스의 사용을 포함한다 - 펼친 손은 이를 켜고, 손 및 팔 방향은 호스를 조준하고, 오므린 손은 이를 끈다.

[0112] 다른 손가락과 손의 인터랙션도 고려된다.

[0113] 도 21a는 타겟 인식, 분석 및 트래킹 시스템에서 사용자의 여러 위치 및 모션을 해석하는 데에 사용되는 컴퓨팅 환경의 일 실시예를 도시한다. 도 1a 내지 도 2와 관련하여 전술한 컴퓨팅 환경(12) 등의 컴퓨팅 환경은 게임 콘솔과 같은 멀티미디어 콘솔(600)일 수 있다. 도 21a에 도시된 바와 같이, 멀티미디어 콘솔(600)은, 레벨 1 캐시(602), 레벨 2 캐시(604) 및 플래시 ROM(606)을 포함하는 CPU(central processing unit, 601)를 갖는다. 레벨 1 캐시(602) 및 레벨 2 캐시(604)는 데이터를 임시로 저장하여 메모리 액세스 사이클의 수를 감소시킴으로써, 처리 속도와 처리율을 향상시킨다. CPU(601)에는 하나 이상의 코어, 따라서 추가적인 레벨 1 및 레벨 2 캐시(602 및 604)가 제공될 수 있다. 플래시 ROM(606)은 멀티미디어 콘솔(600)의 전원이 켜질 때 부팅 프로세스(boot process)의 초기 단계 동안 로딩되는 실행 가능한 코드를 저장할 수 있다.

[0114] 그래픽 처리 장치(GPU, 608)와 비디오 인코더/비디오 코덱(코더/디코더, 614)은 고속 및 고해상도 그래픽 처리를 위해 비디오 처리 파이프라인을 형성한다. 데이터는 버스를 통해 GPU(608)에서 비디오 인코더/비디오 코덱(614)으로 전달된다. 비디오 처리 파이프라인은 텔레비전이나 다른 디스플레이로의 전송을 위해 A/V(오디오/비디오) 포트(640)로 데이터를 출력한다. 메모리 컨트롤러(610)는 GPU(608)에 접속되어, RAM과 같은, 하지만 이에 제한되지는 않는, 다양한 유형의 메모리(612)로의 프로세서 액세스를 용이하게 한다.

[0115] 멀티미디어 콘솔(600)은, 모듈(618)에서 바람직하게 구현되는 I/O 컨트롤러(620), 시스템 관리 컨트롤러(622), 오디오 처리 장치(623), 네트워크 인터페이스 컨트롤러(624), 제 1 USB 호스트 컨트롤러(626), 제 2 USB 호스트 컨트롤러(628), 및 전면 패널 I/O 서브어셈블리(630)를 포함한다. USB 컨트롤러들(626 및 628)은 주변 컨트롤러들(642(1) ~ 642(2)), 무선 어댑터(648)와 외부 메모리 장치(646, 예를 들어, 플래시 메모리, 외부 CD/DVD ROM 드라이브, 이동식 매체 등)를 위한 호스트들로서 기능한다. 네트워크 인터페이스(624) 및/또는 무선 어댑터(648)는 네트워크(예를 들어, 인터넷, 홈 네트워크 등)로의 액세스를 제공하고, 이더넷 카드, 모뎀, 블루투스 모듈, 케이블 모뎀 등을 포함하는 아주 다양한 유무선 어댑터 컴포넌트들 중 임의의 것일 수 있다.

[0116] 시스템 메모리(643)는 부팅 프로세스 동안에 로딩되는 애플리케이션을 저장하기 위해 제공된다. 미디어 드라이브(644)가 제공되고, 이는 DVD/CD 드라이브, 하드 드라이브, 또는 다른 이동식 미디어 드라이브 등을 포함할 수 있다. 미디어 드라이브(644)는 멀티미디어 콘솔(600)의 내부 또는 외부에 있을 수 있다. 애플리케이션 데이터는 멀티미디어 콘솔(600)에 의한 실행, 플레이백 등을 위해 미디어 드라이브(644)를 통해 액세스될 수 있다. 미디어 드라이브(644)는, 직렬 ATA 버스 또는 기타 고속 접속(예를 들어, IEEE 1394)과 같은 버스를 통해 I/O 컨트롤러(620)에 접속된다.

[0117] 시스템 관리 컨트롤러(622)는 멀티미디어 콘솔(600)의 가용성의 보장과 관련된 다양한 서비스 기능들을 제공한다. 오디오 처리 장치(623)와 오디오 코덱(632)은 고신뢰성과 스테레오 처리를 갖는 대용 오디오 처리 파이프라인을 형성한다. 오디오 데이터는 통신 링크를 통해 오디오 처리 장치(623)와 오디오 코덱(632) 사이에서 전달된다. 오디오 처리 파이프라인은 오디오 기능이 있는 외부 오디오 플레이어 또는 장치에 의한 재생을 위해 A/V 포트(640)로 데이터를 출력시킨다.

[0118] 전면 패널 I/O 서브어셈블리(630)는 전원 버튼(650), 배출 버튼(eject button, 652), 및 멀티미디어 콘솔(600)의 외부 표면에 노출된 임의의 LED들(light emitting diodes) 또는 기타 인디케이터들의 기능을 지원한다. 시스템 전원 모듈(636)은 멀티미디어 콘솔(600)의 컴포넌트들에 전력을 공급한다. 팬(fan, 638)은 멀티미디어 콘솔(600) 내의 회로를 냉각시킨다.

[0119] 멀티미디어 콘솔(600) 내의 CPU(601), GPU(608), 메모리 컨트롤러(610) 및 기타 다양한 컴포넌트들은 직렬 및 병렬 버스들, 메모리 버스, 주변장치 버스, 또는 다양한 버스 아키텍처들 중의 임의의 것을 사용하는 프로세서나 로컬 버스를 포함하는 하나 이상의 버스들을 통해 상호접속된다. 예를 들어, 이러한 아키텍처에는 PCI(Peripheral Component Interconnects) 버스, PCI 익스프레스(PCI- Express) 버스 등이 포함될 수 있다.

- [0120] 멀티미디어 콘솔(600)의 전원이 켜질 때, 시스템 메모리(643)로부터 애플리케이션 데이터가 메모리(612) 및/또는 캐시(602, 604)로 로딩되어, CPU(601)에서 실행될 수 있다. 애플리케이션은 멀티미디어 콘솔(600)에서 이용 가능한 다른 미디어 유형들로 네비게이트할 때 일관된 사용자 경험을 제공하는 그래픽 사용자 인터페이스를 제공할 수 있다. 동작 시에는, 미디어 드라이브(644) 내에 포함된 애플리케이션들 및/또는 기타 미디어들이 미디어 드라이브(644)로부터 시작되거나 재생되어, 멀티미디어 콘솔(600)에 추가 기능들을 제공할 수 있다.

[0121] 멀티미디어 콘솔(600)은 텔레비전이나 기타 디스플레이에 시스템을 단순히 접속시킴으로써 독립형 시스템으로서 동작할 수 있다. 이 독립형 모드에서, 멀티미디어 콘솔(600)은 한 명 이상의 사용자들이 시스템과 인터랙트하고, 영화를 보고, 음악을 듣게 하도록 할 수 있다. 그러나, 네트워크 인터페이스(624) 또는 무선 어댑터(648)를 통해 이용 가능하게 된 통합 광대역 접속으로 인해, 멀티미디어 콘솔(600)은 보다 큰 네트워크 커뮤니티의 참가자로서 동작할 수도 있다.

[0122] 멀티미디어 콘솔(600)의 전원이 켜지면, 설정량의 하드웨어 리소스들이 멀티미디어 콘솔 운영 체제에 의한 시스템 사용을 위해 예약된다. 이러한 리소스들은 메모리(예컨대, 16 MB), CPU 및 GPU 사이클(예컨대, 5 %), 네트워킹 대역폭(예컨대, 8 kbs) 등의 예약을 포함할 수 있다. 이러한 리소스들은 시스템 부팅 시간에 예약되기 때문에, 애플리케이션의 관점에서는 예약된 리소스들이 존재하지 않는다.

[0123] 특히, 개시 커널(launch kernel), 동시(concurrent) 시스템 애플리케이션 및 드라이버를 포함할 수 있을 정도로 메모리 예약이 충분히 큰 것이 바람직하다. 예약된 CPU 사용량이 시스템 애플리케이션에서 사용되지 않는 경우, 휴지 쓰레드(idle thread)가 미사용 사이클들을 쓸 수 있도록 CPU 예약이 일정한 것이 바람직하다.

[0124] GPU 예약과 관련하여, 팝업이 오버레이되도록 코드를 스케줄링하는 GPU 인터럽트를 사용하여, 시스템 애플리케이션에서 생성되는 간단한 메시지(예컨대, 팝업)가 디스플레이된다. 오버레이에 필요한 메모리량은 오버레이 영역 크기에 따르며, 오버레이는 스크린 해상도에 맞춰 스케일링되는 것이 바람직하다. 동시 시스템 애플리케이션이 풀 유저 인터페이스(full user interface)를 사용하는 경우에는, 애플리케이션 해상도와는 별개의 해상도를 사용하는 것이 바람직하다. 주파수를 변경하고 TV를 재동기화시킬 필요가 없도록 이 해상도를 설정하기 위해, 스케일러가 사용될 수 있다.

[0125] 멀티미디어 콘솔(600)이 부팅되고 시스템 리소스가 예약된 후에, 동시 시스템 애플리케이션이 실행되어 시스템 기능들을 제공한다. 시스템 기능들은, 전술한 예약된 시스템 리소스들 내에서 실행되는 일련의 시스템 애플리케이션에서 캡슐화되어 있다. 운영 체제 커널은 시스템 애플리케이션 쓰레드인지 게임 애플리케이션 쓰레드인지를 식별한다. 일관적인 시스템 리소스 뷰를 애플리케이션에 제공하기 위해, 시스템 애플리케이션은 사전 설정된 시간 및 간격으로 CPU(601)에서 실행되도록 스케줄링되는 것이 바람직하다. 스케줄링은 콘솔에서 실행되는 게임 애플리케이션에 대한 캐시 중단을 최소화하기 위한 것이다.

[0126] 동시 시스템 애플리케이션이 오디오를 필요로 할 때, 오디오 처리는 시간에 대한 민감도로 인해 게임 애플리케이션과 비동기적으로 스케줄링된다. 시스템 애플리케이션이 동작할 때, 멀티미디어 콘솔 애플리케이션 관리자(이하에서 설명됨)는 게임 애플리케이션 오디오 레벨(예컨대, 음소거(mute), 감쇠(attenuate))을 제어한다.

[0127] 게임 애플리케이션과 시스템 애플리케이션은 입력 장치들(예컨대, 컨트롤러(642(1) 및 642(2)))을 공유한다. 입력 장치들은 예약된 리소스들이 아니지만, 각 시스템 애플리케이션 및 게임 애플리케이션이 입력 장치의 포커스를 갖도록 애플리케이션들 사이에서 스위칭될 것이다. 애플리케이션 관리자는 게임 애플리케이션에 대한 정보 없이 입력 스트림의 스위칭을 제어하는 것이 바람직하며, 드라이버는 포커스 스위치에 관한 상태 정보를 보유한다. 카메라(76, 78) 및 캡쳐 장치(70)는 콘솔(600)을 위한 추가적인 입력 장치들일 수 있다.

[0128] 도 21b는 타겟 인식, 분석 및 트래킹 시스템에서의 여러 위치 및 모션들의 해석, 및/또는 타겟 인식, 분석 및 트래킹 시스템에 의해 디스플레이되는 아바타, 온스크린 캐릭터, 온스크린 객체 등의 가상 캐릭터의 애니메이션화를 위해 사용되는 도 1a 내지 도 2에 도시된 컴퓨팅 환경(12)의 또 다른 실시예인 컴퓨팅 환경(720)을 도시한다. 컴퓨팅 시스템 환경(720)은 적합한 컴퓨팅 환경의 일례일 뿐, 본원에 개시된 대상의 사용이나 기능의 범위를 제한하는 것은 아니다. 컴퓨팅 환경(720)은, 예시적인 운영 환경(720)에 도시된 임의의 컴퓨포넌트 또는 컴퓨포넌트 조합에 관한 의존성이나 요구 사항을 가지는 것으로 해석되어서는 안 된다. 몇몇 실시예에서, 도시된 다양한 컴퓨팅 구성요소들은 본 개시의 특정 양태들에 대한 예를 들어 설명하는 회로를 포함할 수 있다. 예를 들어, 본 개시에서 사용되는 회로는 펌웨어나 스위치로 기능(들)을 실행하도록 구성되는 특화된 하드웨어 컴퓨포넌트들을 포함할 수 있다. 또 다른 실시예에서, 회로는 기능(들)을 실행하는 논리 동작을 구현하는 소프트웨어 명령어들로 구성되는 범용 처리 장치(general purpose processing unit), 메모리 등을 포함할 수 있다. 하드웨어와

소프트웨어의 조합을 포함하는 회로의 실시예들에 있어서, 구현자는 논리를 구현하는 소스 코드를 작성하고, 이 소스 코드는 범용 처리 장치에서 처리될 수 있는 기계 판독 가능 코드로 컴파일될 수 있다. 실시예들에서, 당업자라면 하드웨어, 소프트웨어 또는 하드웨어/소프트웨어의 조합 간에 차이가 거의 없는 방향으로 선행 기술이 발전해왔다는 점을 알 수 있으므로, 특정 기능을 구현하기 위한 하드웨어 대 소프트웨어의 선택은 구현자에게 달려 있는 설계 선택에 불과하다. 특히, 당업자라면 소프트웨어 프로세스가 동등한 하드웨어 구조로 변환가능하고, 하드웨어 구조 또한 동등한 소프트웨어 프로세스로 변환가능하다는 점을 알 수 있다. 따라서, 하드웨어 구현 대 소프트웨어 구현의 선택은 설계 선택으로, 구현자에게 달려 있다.

[0129] 도 21b에서, 컴퓨팅 환경(720)은, 일반적으로 다양한 컴퓨터 판독 가능 매체를 포함하는 컴퓨터(741)를 포함한다. 컴퓨터 판독 가능 매체는 컴퓨터(741)가 액세스할 수 있으며, 휘발성 및 비휘발성 매체, 이동식 및 비이동식 매체 모두를 포함하는 임의의 가용 매체일 수 있다. 시스템 메모리(722)는 ROM(723) 및 RAM(760)과 같은 휘발성 및/또는 비휘발성 메모리의 형태인 컴퓨터 저장 매체를 포함한다. 스타트업 동안과 같이 컴퓨터(741) 내의 구성요소들 사이의 정보 전송을 돋는 기본 루틴을 포함하는 기본 입출력 시스템(BIOS, 724)은 일반적으로 ROM(723)에 저장된다. RAM(760)은 일반적으로 처리 장치(759)에 의해 즉시 액세스 가능 및/또는 바로 동작되는 데이터 및/또는 프로그램 모듈들을 포함한다. 예를 들어, 도 21b는 운영 체제(725), 애플리케이션 프로그램(726), 다른 프로그램 모듈(727), 및 프로그램 데이터(728)를 도시하고 있으며, 이들로서 제한되는 것은 아니다. 도 21b는 고속 및 고해상도 그래픽 처리 및 저장을 위한 관련 비디오 메모리(730)를 갖고 있는 그래픽 처리 장치(GPU, 729)도 포함한다. GPU(729)는 그래픽 인터페이스(731)를 통해 시스템 버스(721)에 접속될 수 있다.

[0130] 또한, 컴퓨터(741)는 다른 이동식/비이동식, 휘발성/비휘발성 컴퓨터 저장 매체를 포함할 수 있다. 예를 들어, 도 21b는 비이동식, 비휘발성 자기 매체에 판독 또는 기록하는 하드 디스크 드라이브(738), 이동식, 비휘발성 자기 디스크(754)에 판독 또는 기록하는 자기 디스크 드라이브(739), 및 CD ROM 또는 다른 광학 매체와 같은 이동식, 비휘발성 광 디스크(753)에 판독 또는 기록하는 광 디스크 드라이브(740)를 도시한다. 예시적인 운영 환경에서 사용 가능한 다른 이동식/비이동식, 휘발성/비휘발성 컴퓨터 저장 매체는, 자기 테이프 카세트, 플래시 메모리 카드, DVD, 디지털 비디오 테이프, 고체 상태 RAM, 고체 상태 ROM 등을 포함하지만, 이들로서 제한되는 것은 아니다. 하드 디스크 드라이브(738)는 일반적으로 인터페이스(734)와 같은 비이동식 메모리 인터페이스를 통해 시스템 버스(721)에 연결되며, 자기 디스크 드라이브(739) 및 광 디스크 드라이브(740)는 일반적으로 인터페이스(735)와 같은 이동식 메모리 인터페이스에 의해 시스템 버스(721)에 연결된다.

[0131] 앞서 논의되었으며 도 21b에 도시된 드라이브 및 그 관련 컴퓨터 저장 매체는 컴퓨터(741)를 위한 컴퓨터 판독 가능 명령어, 데이터 구조, 프로그램 모듈 및 기타 데이터를 저장한다. 도 21b에서, 예를 들어, 하드 디스크 드라이브(738)는 운영 체제(758), 애플리케이션 프로그램(757), 다른 프로그램 모듈(756) 및 프로그램 데이터(755)를 저장하는 것으로 도시되어 있다. 이러한 컴퓨터들은 운영 체제(725), 애플리케이션 프로그램(726), 다른 프로그램 모듈(727) 및 프로그램 데이터(728)와 동일할 수도 또는 다를 수도 있음을 알 것이다. 운영 체제(758), 애플리케이션 프로그램(757), 다른 프로그램 모듈(756), 및 프로그램 데이터(755)는, 최소한 이들이 상이한 카피들임을 나타내기 위해 본원에서 상이한 번호가 부여된다. 사용자는 키보드(751), 및 일반적으로 마우스, 트랙볼, 또는 터치 패드로 불리는 포인팅 장치(752)를 비롯한 입력 장치들을 통해 명령 및 정보를 컴퓨터(741)에 입력할 수 있다. 다른 입력 장치들(도시되지 않음)은 마이크, 조이스틱, 게임 패드, 위성 접시, 스캐너, 리모콘 등을 포함할 수 있다. 이들 및 다른 입력 장치들은 보통 시스템 버스에 연결된 사용자 입력 인터페이스(736)를 통해 처리 장치(759)에 접속되지만, 별도로 다른 인터페이스와 버스 구조에 의해 접속될 수도 있다. 카메라(26, 28) 및 캡쳐 장치(20)는 콘솔(700)의 추가 입력 장치에 해당한다. 모니터(742) 또는 다른 형태의 디스플레이 장치도 비디오 인터페이스(732)와 같은 인터페이스를 통해 시스템 버스(721)에 접속된다. 모니터뿐만 아니라, 컴퓨터들은, 출력 주변 기기 인터페이스(733)를 통해 접속될 수 있는 스피커(744) 및 프린터(743)를 비롯한 다른 주변 기기 출력 장치들을 포함할 수 있다.

[0132] 컴퓨터(741)는 원격 컴퓨터(746)와 같은 하나 이상의 원격 컴퓨터들에 대한 논리적인 연결들을 사용하여 네트워킹된 환경에서 동작할 수 있다. 원격 컴퓨터(746)는 개인용 컴퓨터, 서버, 라우터, 네트워크 PC, 피어 장치(peer device) 또는 다른 공통 네트워크 노드일 수 있으며, 단지 메모리 저장 장치(747)만이 도 21b에 도시되어 있지만, 일반적으로 컴퓨터(741)와 관련하여 전술한 다수의 또는 모든 구성요소들을 포함한다. 도 21b에 도시된 논리적인 연결은 LAN(local area network, 745) 및 WAN(wide area network, 749)을 포함하지만, 또한 다른 네트워크들도 포함할 수 있다. 이러한 네트워킹 환경들은 사무실, 기업(enterprise-wide) 컴퓨터 네트워크, 인트

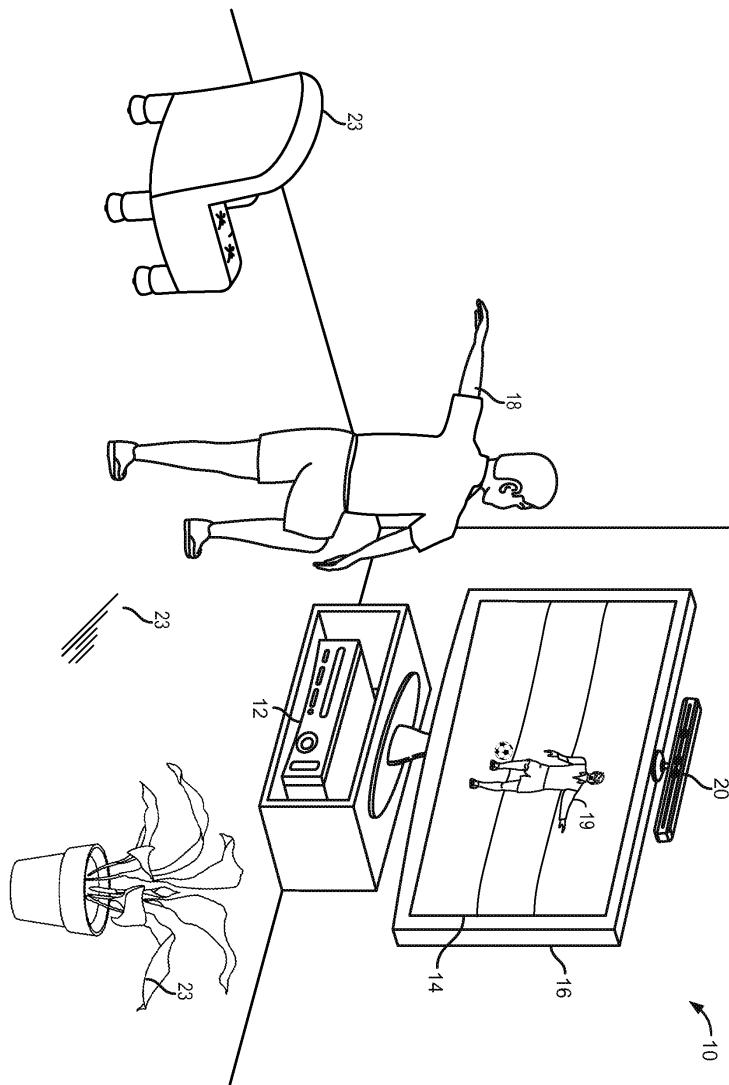
라넷 및 인터넷에 흔하다.

[0133] LAN 네트워킹 환경에서 사용될 때, 컴퓨터(741)는 네트워크 인터페이스 또는 어댑터(737)를 통해서 LAN(745)에 연결된다. WAN 네트워킹 환경에서 사용될 때, 컴퓨터(741)는 일반적으로 모뎀(750) 또는 인터넷과 같이 WAN(749)을 통해 통신을 구축하기 위한 다른 수단을 포함한다. 내부 또는 외부에 존재할 수 있는 모뎀(750)은 사용자 입력 인터페이스(736), 또는 다른 적절한 메카니즘을 통해 시스템 버스(721)에 연결될 수 있다. 네트워킹 환경에서, 컴퓨터(741) 또는 그 일부분에 관련하여 도시된 프로그램 모듈들이 원격 메모리 저장 장치 내에 저장될 수 있다. 예를 들면, 도 21b는 메모리 장치(747) 상에 존재하는 원격 애플리케이션 프로그램들(748)을 도시하고 있다. 도시된 네트워크 연결들은 예시적인 것이며, 컴퓨터들 간에 통신 링크를 구축하는 다른 수단이 사용될 수 있음을 이해할 것이다.

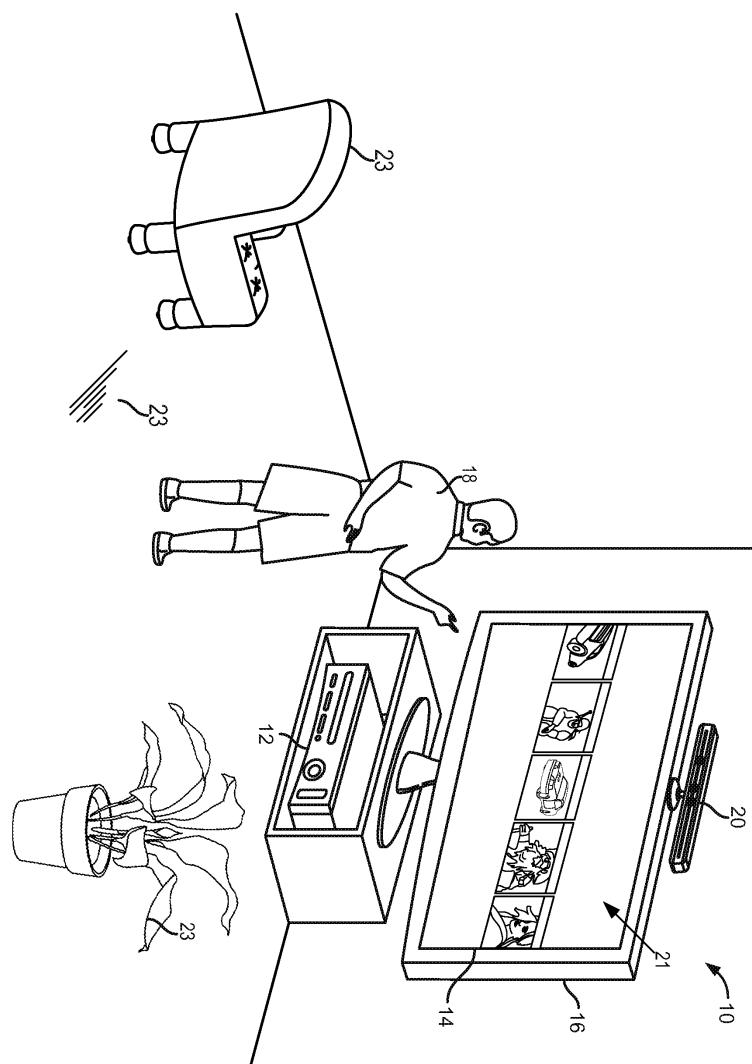
[0134] 본 발명의 시스템의 전술된 상세한 설명은 예시와 설명의 목적으로 제시되었다. 이는 포괄적이거나 본 발명의 시스템을 개시된 정확한 형태로 제한하기 위한 것이 아니다. 많은 수정과 변형이 전술된 개시에 기초하여 가능하다. 설명된 실시예는 본 발명의 시스템의 원리 및 그의 실질적인 애플리케이션을 가장 잘 설명함으로써 당업자들이 고려되는 특정 사용에 적합하게 된 바와 같이, 다양한 실시예에서 그리고 다양한 수정과 함께 본 발명의 시스템을 가장 잘 이용하게 하기 위해 선택되었다. 본 발명의 시스템의 범위는 본 명세서에 첨부된 특허청구범위에 의해 제한되도록 의도된다.

도면

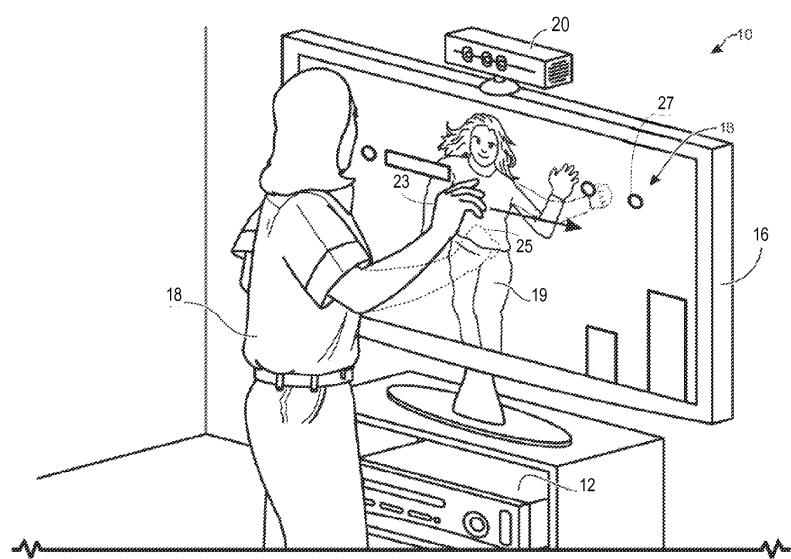
도면1a



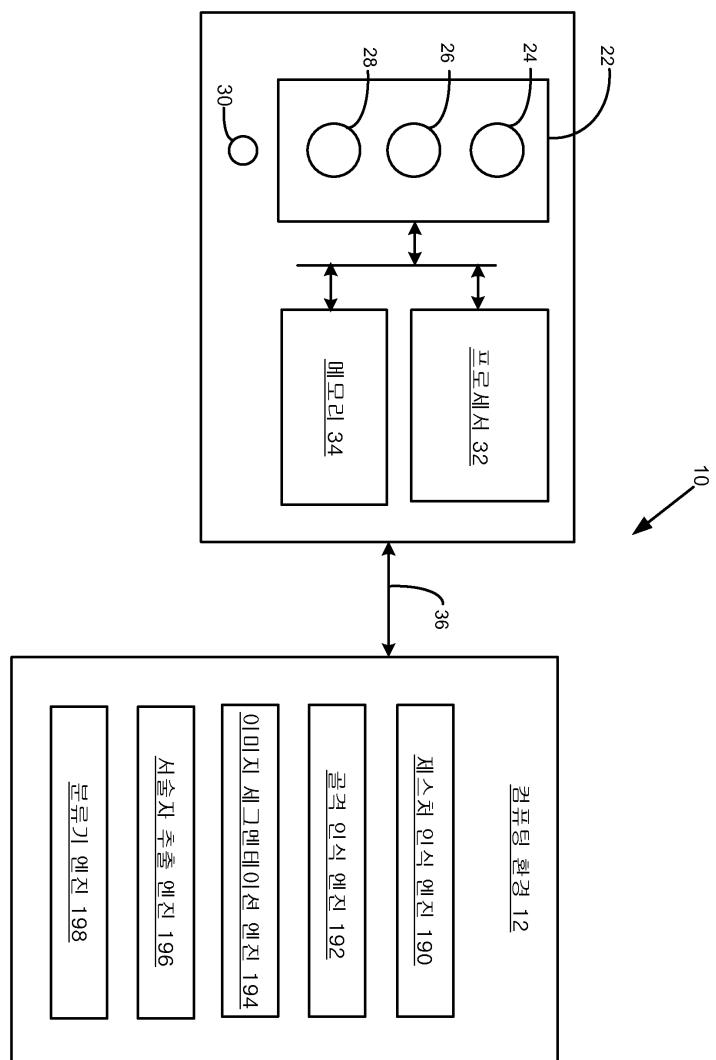
도면1b



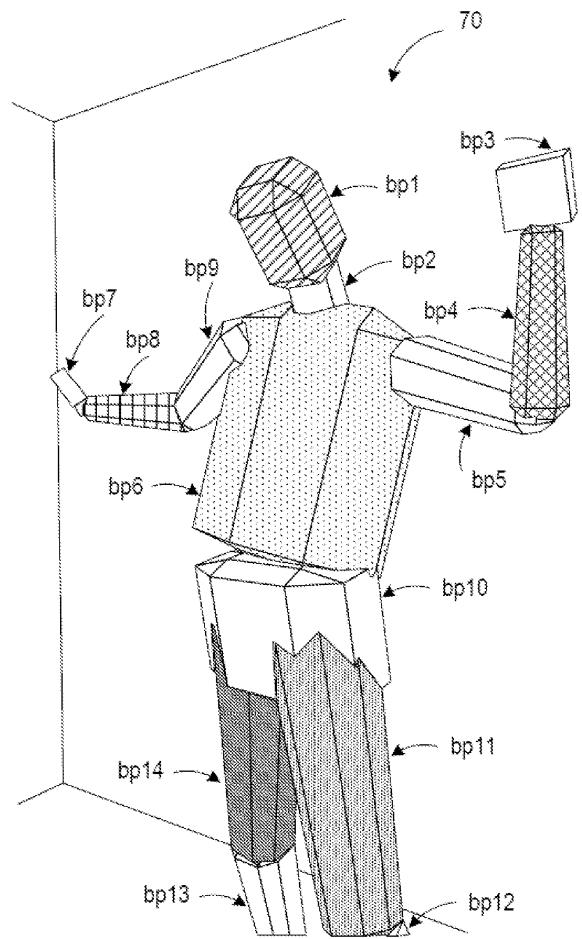
도면1c



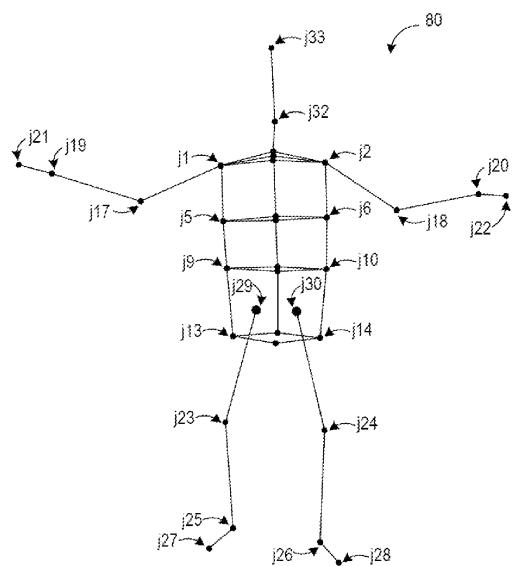
도면2



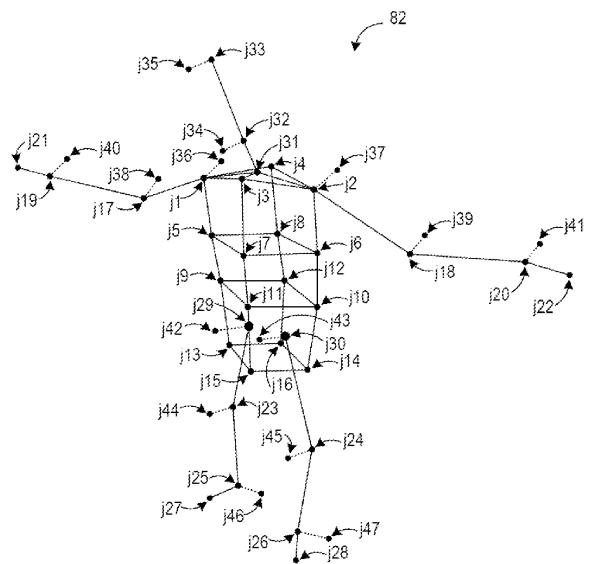
도면3



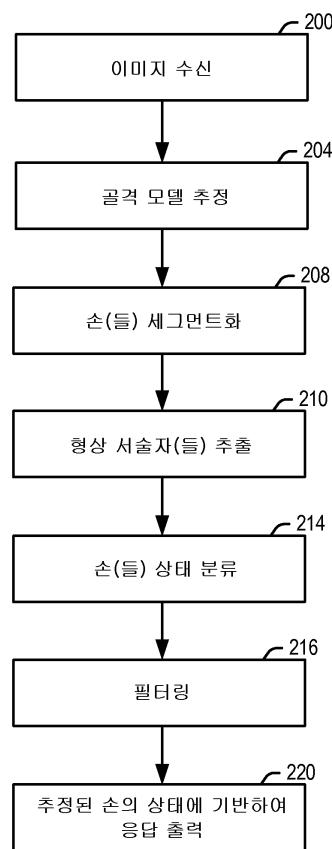
도면4



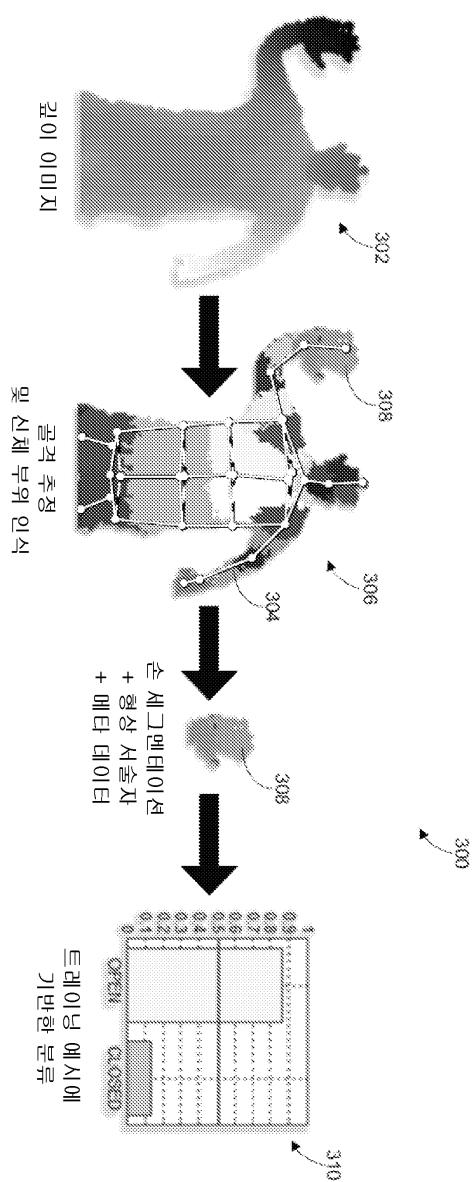
도면5



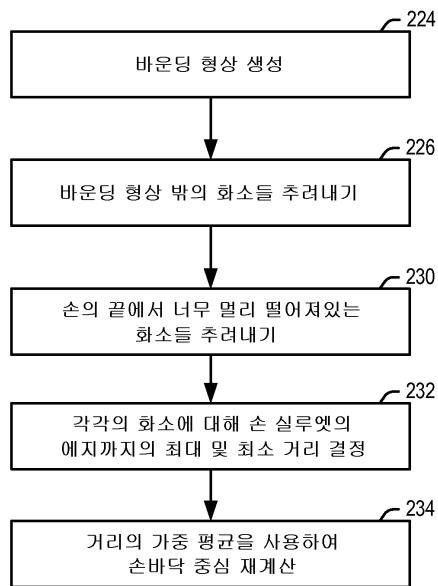
도면6



도면7



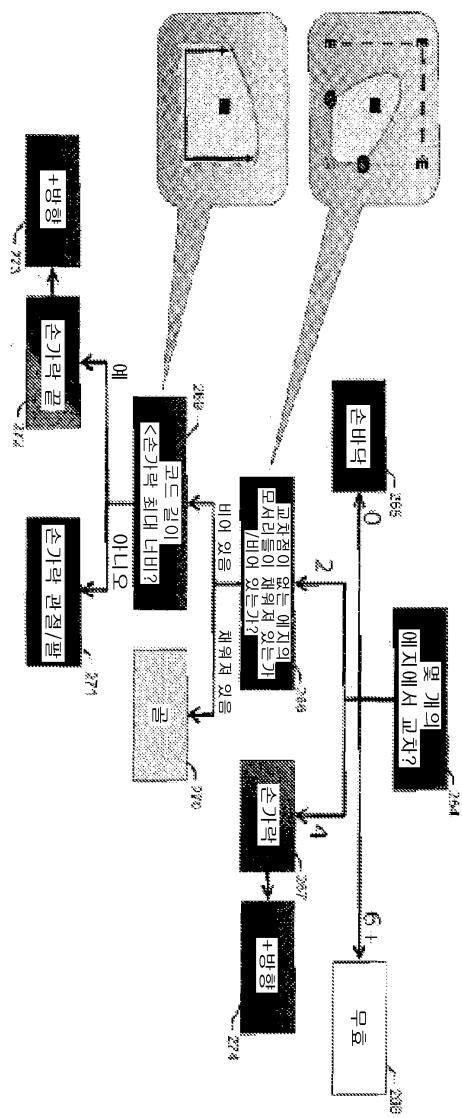
도면8



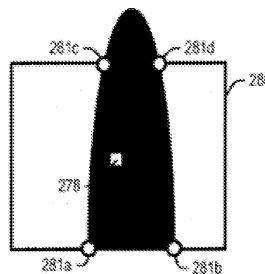
도면9



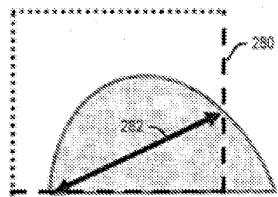
도면10



도면12



도면13

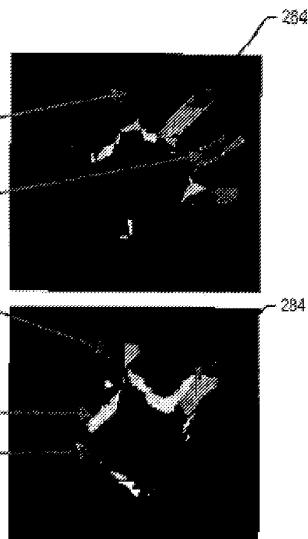


도면14

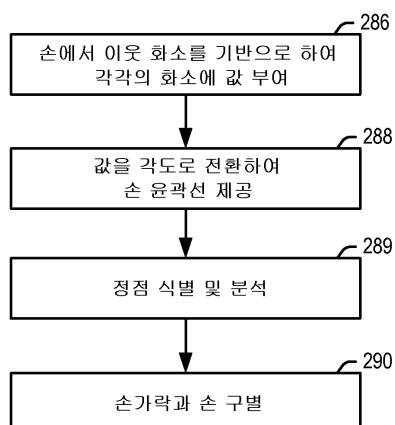
화소 분류기

결과

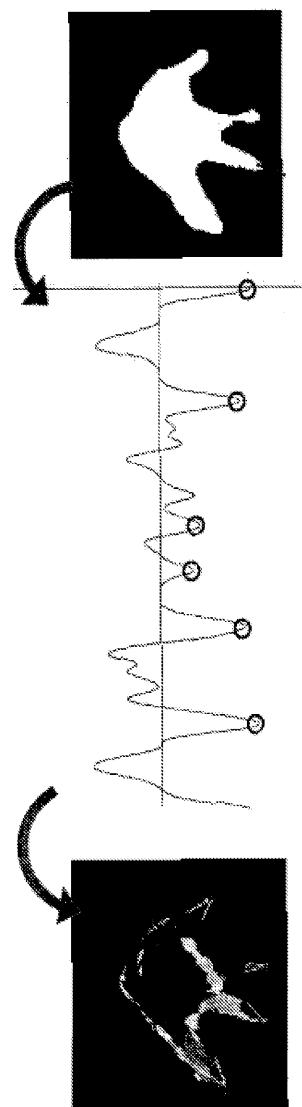
- 손가락 끝 + 방향
- 손가락 지골 + 방향
- 더 큰 그룹(2/4 손가락)
 - 더 큰 경계 반경으로 다시 분류기 실행
- 손가락 베이스
- 손바닥



도면15

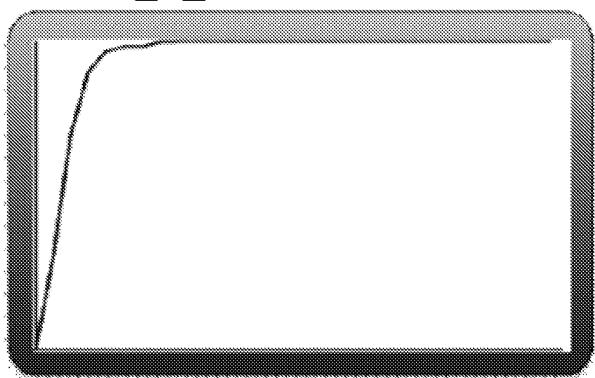


도면16

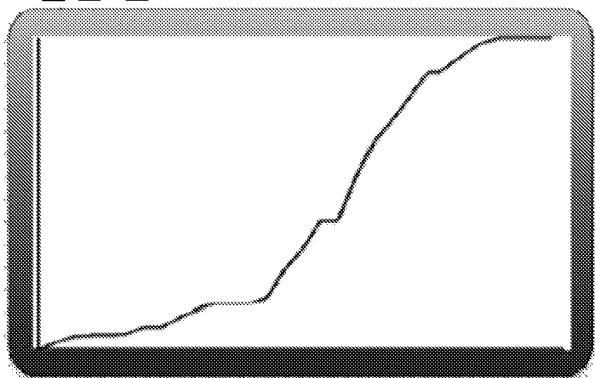


도면17

오므린 손

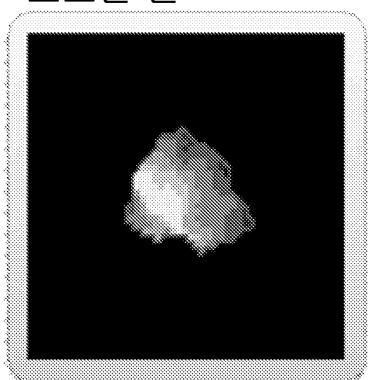


펼친 손

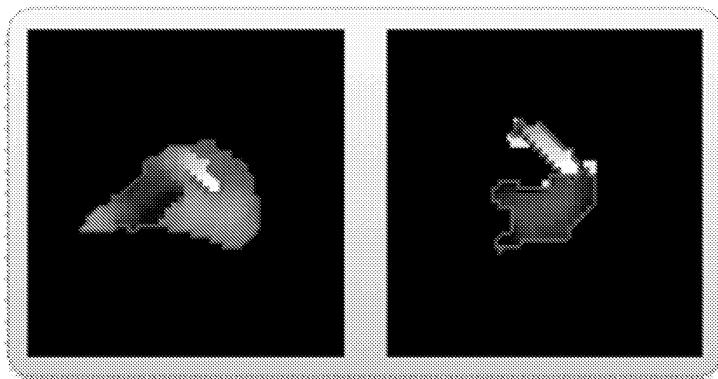


도면18

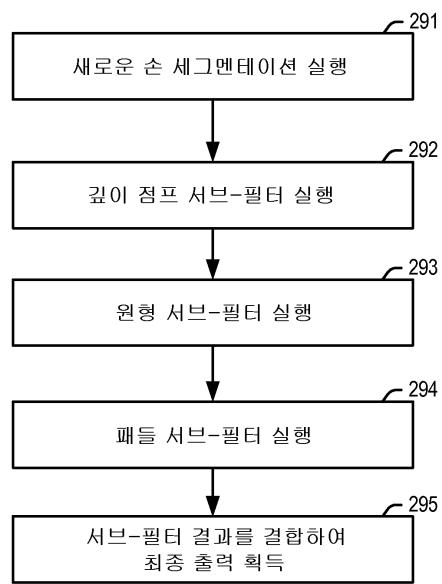
오므린 손



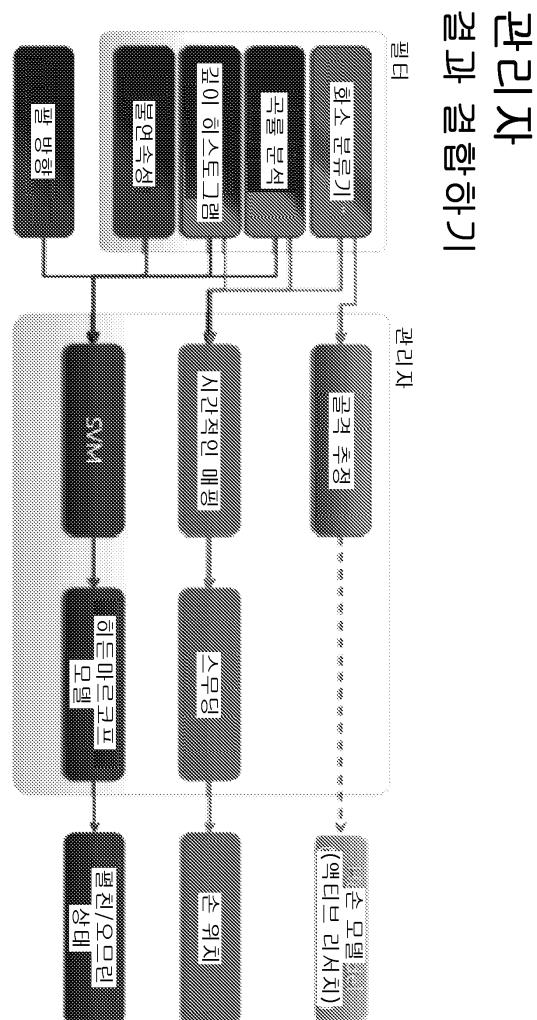
펼친 손



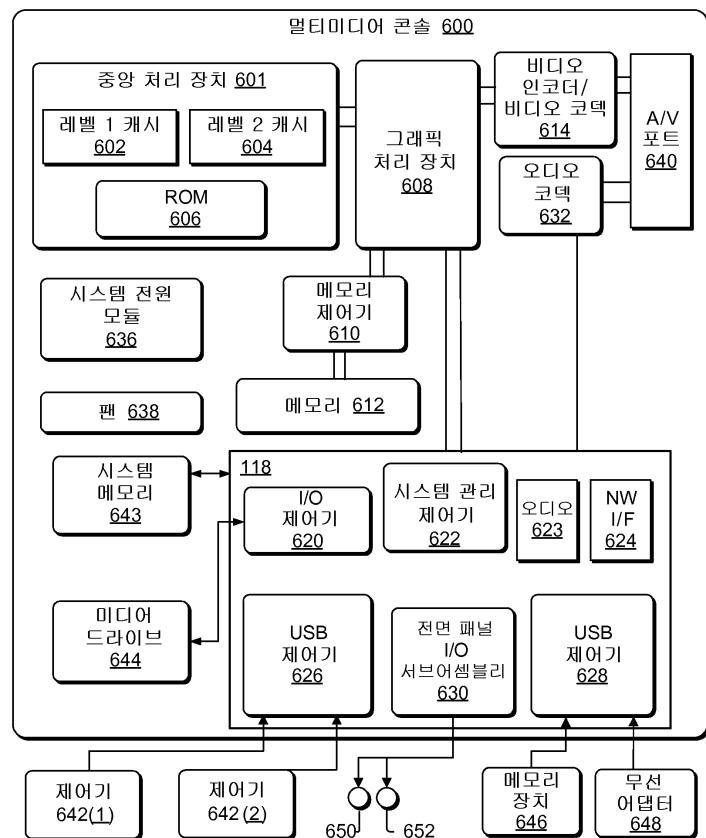
도면19



도면20



도면21a



도면21b

