



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2014-0019765
(43) 공개일자 2014년02월17일

(51) 국제특허분류(Int. Cl.)
H04N 13/02 (2006.01) H04N 5/225 (2006.01)
G03B 21/20 (2006.01) G03B 21/14 (2006.01)
(21) 출원번호 10-2013-7005894
(22) 출원일자(국제) 2011년08월01일
심사청구일자 없음
(85) 번역문제출일자 2013년03월07일
(86) 국제출원번호 PCT/US2011/046139
(87) 국제공개번호 WO 2012/033578
국제공개일자 2012년03월15일
(30) 우선권주장
12/877,595 2010년09월08일 미국(US)

(71) 출원인
마이크로소프트 코포레이션
미국 워싱턴주 (우편번호 : 98052) 레드몬드 원
마이크로소프트 웨이
(72) 발명자
카츠 사기
미국 워싱턴주 98052-6399 레드몬드 원 마이크로
소프트 웨이 엘씨에이 - 인터내셔널 페이턴츠 마
이크로소프트 코포레이션
에들러 아비샤이
미국 워싱턴주 98052-6399 레드몬드 원 마이크로
소프트 웨이 엘씨에이 - 인터내셔널 페이턴츠 마
이크로소프트 코포레이션
(74) 대리인
제일특허법인

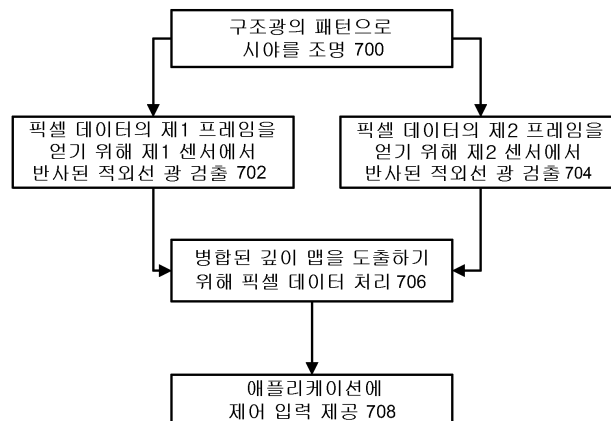
전체 청구항 수 : 총 15 항

(54) 발명의 명칭 구조광 및 스테레오 비전에 기초한 깊이 카메라

(57) 요약

시야 내 사용자의 모션을 추적하는 시스템에서와 같이, 깊이 카메라 시스템은 구조광 조명기와 적외선 광 검출기와 같은 다수의 센서를 이용한다. 하나의 센서는 짧은 범위 검출에 대해 최적화될 수 있는 반면 다른 센서는 긴 범위 검출에 대해 최적화된다. 센서들은 조명기로부터의 상이한 베이스라인 거리뿐만 아니라, 상이한 공간 해상도, 노출 시간 및 감도를 가질 수 있다. 다른 접근에서, 깊이 값은 구조광 패턴에 매칭시킴으로써 각각의 센서로부터 획득되고, 애플리케이션에 대한 입력으로 제공되는 최종 깊이 맵을 획득하기 위해 깊이 값이 병합된다. 병합에는, 비가중 평균, 가중 평균, 정확도 척도 및/또는 신뢰 척도가 관여될 수 있다. 다른 접근에서, 병합에 포함되는 추가 깊이 값은 센서의 픽셀 데이터 사이의 스테레오스코픽 매칭을 이용하여 획득된다.

대표도



특허청구의 범위

청구항 1

구조광(structured light)의 패턴으로 시야 내 물체를 조명하는 조명기와,

픽셀 데이터의 제1 프레임을 획득하기 위해 상기 물체로부터 반사된 광을 감지하는 제1 센서-상기 제1 센서는 더 짧은 범위 이미징에 최적화된-와,

픽셀 데이터의 제2 프레임을 획득하기 위해 상기 물체로부터 반사된 광을 감지하는 제2 센서-상기 제2 센서는 더 긴 범위 이미징에 최적화된-와,

픽셀 데이터의 상기 제1 프레임을 상기 구조광의 상기 패턴과 비교함으로써 상기 물체의 제1 구조광 깊이 맵을 도출하고, 픽셀 데이터의 상기 제2 프레임을 상기 구조광의 상기 패턴과 비교함으로써 상기 물체의 제2 구조광 깊이 맵을 도출하고, 상기 제1 및 제2 구조광 깊이 맵에 기초하는 병합된 깊이 맵을 도출하는, 적어도 하나의 제어 회로를 포함하는

깊이 카메라 시스템.

청구항 2

제 1 항에 있어서,

상기 제1 센서와 상기 조명기 사이의 베이스라인 거리(BL1)는 상기 제2 센서와 상기 조명기 사이의 베이스라인 거리(BL2)보다 작은

깊이 카메라 시스템.

청구항 3

제 2 항에 있어서,

상기 제1 센서의 노출 시간은 상기 제2 센서의 노출 시간보다 짧은

깊이 카메라 시스템.

청구항 4

제 2 항에 있어서,

상기 제1 센서의 감도는 상기 제2 센서의 감도보다 작은

깊이 카메라 시스템.

청구항 5

제 2 항에 있어서,

상기 제1 센서의 공간 해상도는 상기 제2 센서의 공간 해상도보다 작은

깊이 카메라 시스템.

청구항 6

제 1 항에 있어서,

상기 제2 구조광 깊이 맵은 깊이 값을 포함하고,

상기 병합된 깊이 맵을 도출하는데 있어서, 상기 제2 구조광 깊이 맵의 상기 깊이 값은 상기 제1 구조광 깊이 맵의 깊이 값보다 더 강하게 가중되는

깊이 카메라 시스템.

청구항 7

제 1 항에 있어서,

상기 적어도 하나의 제어 회로는 상기 물체의 적어도 하나의 스테레오스코픽 깊이 맵에 기초하여 상기 병합된 깊이 맵을 도출하고,

상기 적어도 하나의 제어 회로는, (i)픽셀 데이터의 상기 제2 프레임에 픽셀 데이터의 상기 제1 프레임을 스테레오스코픽 매칭하는 것과 (ii)픽셀 데이터의 상기 제1 프레임에 픽셀 데이터의 상기 제2 프레임을 스테레오스코픽 매칭하는 것 중 적어도 하나에 의해 상기 적어도 하나의 스테레오스코픽 깊이 맵을 도출하는

깊이 카메라 시스템.

청구항 8

제 7 항에 있어서,

상기 제1 및 제2 구조광 깊이 맵과 상기 적어도 하나의 스테레오스코픽 깊이 맵은 깊이 값을 포함하고,

상기 적어도 하나의 제어 회로는 상기 물체의 상기 제1 구조광 깊이 맵의 상기 깊이 값에 제1 세트의 가중치를 할당하고, 상기 물체의 상기 제2 구조광 깊이 맵의 상기 깊이 값에 제2 세트의 가중치를 할당하며, 상기 물체의 상기 스테레오스코픽 깊이 맵의 상기 깊이 값에 제3 세트의 가중치를 할당하고, 상기 제1, 제2 및 제3 세트의 가중치에 기초하여 상기 병합된 깊이 맵을 도출하는

깊이 카메라 시스템.

청구항 9

제 8 항에 있어서,

상기 제1 세트의 가중치는 상기 제1 센서와 상기 조명기 사이의 베이스라인 거리에 기초하여 할당되고,

상기 제2 세트의 가중치는 상기 제2 센서와 상기 조명기 사이의 베이스라인 거리에 기초하여 할당되며,

상기 제3 세트의 가중치는 상기 제1 및 제2 센서 사이의 베이스라인 거리에 기초하여 할당되는

깊이 카메라 시스템.

청구항 10

깊이 카메라 시스템에서 이미지 데이터를 처리하는 방법으로서,

구조광의 패턴으로 시야 내 물체를 조명하는 단계와,

제1 센서에서, 픽셀 데이터의 제1 프레임을 획득하기 위해 상기 물체로부터 반사된 광을 감지하는 단계와,

제2 센서에서, 픽셀 데이터의 제2 프레임을 획득하기 위해 상기 물체로부터 반사된 광을 감지하는 단계와,

픽셀 데이터의 상기 제1 프레임을 상기 구조광의 상기 패턴과 비교함으로써 상기 물체의 제1 구조광 깊이 맵을 도출하는 단계-상기 제1 구조광 깊이 맵은 픽셀 데이터의 상기 제1 프레임의 픽셀에 대한 깊이 값을 포함함-와,

픽셀 데이터의 상기 제2 프레임을 상기 구조광의 상기 패턴과 비교함으로써 상기 물체의 제2 구조광 깊이 맵을 도출하는 단계-상기 제2 구조광 깊이 맵은 픽셀 데이터의 상기 제2 프레임의 픽셀에 대한 깊이 값을 포함-와,
픽셀 데이터 맵의 상기 제1 프레임의 하나 이상의 픽셀의 상기 깊이 값의 개선(refinement)이 희망되는지 여부를 결정하는 단계와,
상기 개선이 희망되는 경우, 픽셀 데이터의 상기 제1 프레임의 상기 하나 이상의 픽셀을 픽셀 데이터의 상기 제2 프레임의 하나 이상의 픽셀에 스테레오스코픽 매칭하는 것을 수행하는 단계를 포함하는
이미지 데이터 처리 방법.

청구항 11

제 10 항에 있어서,
픽셀 데이터의 상기 제1 프레임을 상기 구조광의 상기 패턴과 비교하는데 있어 픽셀 데이터의 상기 제1 프레임의 상기 하나 이상의 픽셀이 구조광의 상기 패턴에 성공적으로 매칭되지 않은 때에 상기 개선이 희망되는
이미지 데이터 처리 방법.

청구항 12

제 10 항에 있어서,
상기 개선은 상기 깊이 값이 문턱 거리를 초과하는 때에 희망되는
이미지 데이터 처리 방법.

청구항 13

제 10 항에 있어서,
상기 제1 및 제2 센서 사이의 베이스라인 거리($BL1+BL2$)는 상기 제1 센서와 조명기 사이의 베이스라인 거리($BL1$)보다 크고, 상기 제2 센서와 상기 조명기 사이의 베이스라인 거리($BL2$)보다 큰
이미지 데이터 처리 방법.

청구항 14

제 10 항에 있어서,
상기 스테레오스코픽 매칭은 상기 개선이 희망되는 픽셀 데이터의 상기 제1 프레임의 상기 하나 이상의 픽셀에 대해 수행되지만, 상기 개선이 희망되지 않는 픽셀 데이터의 상기 제1 프레임의 하나 이상의 다른 픽셀에 대해서는 수행되지 않는
이미지 데이터 처리 방법.

청구항 15

제 10 항에 있어서,
상기 개선이 희망되는 경우, 상기 스테레오스코픽 매칭 및 상기 제1 및 제2 구조광 깊이 맵에 기초하여 병합 깊이 맵을 제공하는 단계를 더 포함하는
이미지 데이터 처리 방법.

명세서

배경 기술

- [0001] 실시간 깊이 카메라(depth camera)는 카메라의 시야에서 사람이나 다른 물체까지의 거리를 결정하고, 카메라의 프레임 레이트에 기초하여 실질적으로 실시간으로 거리를 업데이트할 수 있다. 이러한 깊이 카메라는, 예를 들어, 물리적인 공간에서 인체나 기타 객체의 위치와 움직임에 관한 데이터를 획득하기 위한 모션 캡처 시스템에서 사용될 수 있고, 이 데이터를 연산 시스템에서 애플리케이션의 입력으로 사용할 수 있다. 군용, 엔터테인먼트용, 스포츠용 및 의료용과 같은 많은 애플리케이션이 가능하다. 통상, 깊이 카메라는 시야를 비추는 조명과 이미지를 형성하기 위해 시야로부터 광을 감지하는 이미지 센서를 포함한다. 그러나, 광 조건, 표면 질감(texture)과 색, 그리고 오클루전(occlusion)의 가능성과 같은 변수 때문에 다양한 문제가 존재한다.

발명의 내용

과제의 해결 수단

- [0002] 깊이 카메라 시스템이 제공된다. 실질적으로 실시간으로 깊이 맵을 획득하기 위해, 깊이 카메라 시스템은 적어도 2개의 이미지 센서와, 구조광 이미지 처리와 스테레오스코픽 이미지 처리의 조합을 사용한다. 깊이 맵은 센서에 의해 획득되는 픽셀 데이터의 새로운 프레임 각각에 대해 업데이트될 수 있다. 또한, 조명기로부터 상이한 거리에 이미지 센서가 설치될 수 있고, 상이한 특성을 가질 수 있어, 오클루전의 확률을 감소시키면서 더 정확한 깊이 맵이 획득될 수 있도록 한다.
- [0003] 일 실시형태에서, 깊이 카메라 시스템은 구조광의 패턴으로 시야 내 물체를 조명하는 조명기, 적어도 제1 및 제2 센서 및 적어도 하나의 제어 회로를 포함한다. 제1 센서는 픽셀의 제1 프레임을 획득하기 위해 물체로부터 반사된 광을 감지하고, 짧은 범위 이미징에 최적화된다. 이 최적화는, 예를 들어, 제1 센서와 조명기 사이의 상대적으로 짧은 거리, 또는 제1 센서의 상대적으로 작은 노출 시간, 공간 해상도 및/또는 광에 대한 민감도의 관점에서 실현될 수 있다. 깊이 카메라 시스템은 픽셀 데이터의 제2 프레임을 획득하기 위해 물체로부터 반사된 광을 감지하는 제2 센서를 더 포함하는데, 제2 센서는 긴 범위 이미징에 최적화된다. 이 최적화는, 예를 들어, 제2 센서와 조명기 사이의 상대적으로 긴 거리, 또는 제2 센서의 상대적으로 큰 노출 시간, 공간 해상도 및/또는 광에 대한 민감도의 관점에서 실현될 수 있다.
- [0004] 깊이 카메라 시스템은 적어도 하나의 제어 회로를 더 포함하는데, 이는 센서 및 조명기와 공통 하우징 내에 있을 수 있고 및/또는 컴퓨팅 환경과 같은 별도의 컴포넌트에 있을 수 있다. 적어도 하나의 제어 회로는 픽셀 데이터의 제1 프레임을 구조광의 패턴과 비교함으로써 물체의 제1 구조광 깊이 맵을 도출하고, 픽셀 데이터의 제2 프레임을 구조광의 패턴과 비교함으로써 물체의 제2 구조광 깊이 맵을 도출하고, 제1 및 제2 구조광 깊이 맵에 기초하는 병합된 깊이 맵을 도출한다. 각각의 깊이 맵은 픽셀의 그리드에서와 같이 각 픽셀 위치에 대한 깊이 값을 포함한다.
- [0005] 다른 태양에서, 스테레오스코픽 이미지 처리는 깊이 값을 개선하는데도 사용된다. 예를 들어, 픽셀 데이터의 제1 및/또는 제2 프레임의 하나 이상의 픽셀이 구조광의 패턴에 성공적으로 매칭되지 않은 때, 또는 깊이 값이 양호한 정확도를 달성하기 위해 더 큰 베이스라인을 요구하는 더 큰 거리를 나타내는 때에 스테레오스코픽 이미지 처리의 사용이 트리거될 수 있다. 이러한 방식으로, 필요에 따라서만 깊이 값에 대해 추가의 개선이 제공되고, 불필요한 처리 단계를 회피한다.
- [0006] 일부 경우에, 센서에 의해 획득되는 깊이 데이터에는, 센서의 특성 및/또는 깊이 값의 신뢰도에 기초하는 정확도 척도에 기초하여 가중치가 할당될 수 있다.
- [0007] 최종 깊이 맵은 예를 들어, 모션 캡처 시스템에서 애플리케이션에 대한 입력으로 사용될 수 있는데, 모션 캡처 시스템에 의해 추적되는 물건은 사람이고, 애플리케이션은, 아바타를 움직이거나, 온-스크린 메뉴를 탐색(navigate)하거나, 어떤 다른 동작을 수행함으로써, 사람에 의한 제스처 또는 움직임에 응답하여 모션 캡처 시스템의 디스플레이를 변경한다.
- [0008] 본 개요는 아래의 설명에서 더 설명될 개념 중 선택된 것을 단순화된 형태로 소개하기 위하여 제공된다. 본 개요는 청구된 주제의 핵심 특징 또는 중요 특징을 식별하려는 것이 아니고, 청구된 주제의 범위를 제한하는데 사

용되려는 것도 아니다.

도면의 간단한 설명

[0009]

도면에서, 유사한 번호가 주어진 구성요소는 서로 대응한다.

도 1은 모션 캡처 시스템의 예시적인 실시형태를 도시한다.

도 2는 도 1의 모션 캡처 시스템의 예시적인 블록도를 도시한다.

도 3은 도 1의 모션 캡처 시스템에서 사용될 수 있는 컴퓨팅 환경의 예시적인 블록도를 도시한다.

도 4는 도 1의 모션 캡처 시스템에서 사용될 수 있는 컴퓨팅 환경의 다른 예시적인 블록도를 도시한다.

도 5(a)는 구조광 시스템에서 캡처된 프레임과 광 프레임을 도시한다.

도 5(b)는 스테레오스코픽 광 시스템에서 2개의 캡처된 프레임을 도시한다.

도 6(a)는 조명기의 공통측에 2개의 센서를 갖는 이미징 컴포넌트를 도시한다.

도 6(b)는 조명기의 일 측에 2개의 센서를 갖고, 조명기의 반대측에 하나의 센서를 갖는 이미징 컴포넌트를 도시한다.

도 6(c)는 조명기의 공통측에 3개의 센서를 갖는 이미징 컴포넌트를 도시한다.

도 6(d)는 조명기의 반대측에 2개의 센서를 갖는 이미징 컴포넌트를 도시하여, 2개의 센서가 물체의 상이한 부분을 어떻게 감지하는지를 보여준다.

도 7(a)는 시야의 깊이 맵을 얻기 위한 프로세스를 도시한다.

도 7(b)는 2개의 구조광 깊이 맵이 병합(merge)되는 도 7(a)의 단계 706을 더 자세히 도시한다.

도 7(c)는 2개의 구조광 깊이 맵과 2개의 스테레오스코픽 깊이 맵이 병합되는, 도 7(a)의 단계 706을 더 자세히 도시한다.

도 7(d)는 필요에 따라 스테레오스코픽 매칭을 이용하여 깊이 값이 개선되는, 도 7(a)의 단계 706을 더 자세히 도시한다.

도 7(e)는 필요에 따라 스테레오스코픽 매칭을 이용하여 병합된 깊이 맵의 깊이 값이 개선되는, 도 7(a)의 단계 706에 대한 다른 접근을 더 상세히 도시한다.

도 8은 도 7(a)의 단계 708에 제시된 바와 같이 제어 입력을 이용하여 사람 타겟을 추적하기 위한 예시적인 방법을 도시한다.

도 9는 도 8의 단계 808에 제시된 바와 같이 사람 타겟의 예시적인 모델을 도시한다.

발명을 실시하기 위한 구체적인 내용

[0010]

깊이 카메라는 시야 내 하나 이상의 물체를 추적하는데 사용하기 위해 제공된다. 예시적인 구현에서, 깊이 카메라는 사용자를 추적하기 위해 모션 트래킹(motion tracking) 시스템에서 사용된다. 깊이 카메라는 광 조건, 표면 질감과 색, 그리고 오클루전의 가능성과 같은 주소 변수에 최적화되는 2 이상의 센서를 포함한다. 최적화는 센서의 배치를 서로에 대해 그리고 조명기에 대해 최적화하는 것뿐만 아니라 센서의 노출 시간, 감도 및 공간 해상도를 최적화하는 것도 포함한다. 최적화는 또한, 픽셀 데이터의 프레임을 구조광(structured light)의 패턴에 매칭시키는 것 및/또는 픽셀 데이터의 프레임을 다른 프레임에 매칭시키는 것 등에 의해 깊이 맵 데이터가 획득되는 방법을 최적화하는 것을 포함할 수 있다.

[0011]

여기에 도시된 바와 같은 다수 센서의 사용은 다른 접근에 비해 장점을 제공한다. 예를 들어, 다른 스테레오 카메라보다 실시간 깊이 카메라는 2-D 매트릭스에 임베딩될 수 있는 깊이 맵을 제공하는 경향이 있다. 이러한 카메라는 보통 깊이 맵을 추출하기 위해 단일의 이미징 장치를 사용하고, 따라서 오클루전된 객체에 대해서는 아무런 정보도 주어지지 않기 때문에, 이러한 카메라를 때론 2.5D 카메라라고도 한다. 스테레오 깊이 카메라는 2 이상의 센서에 의해 보일 수 있는 위치의 다소 성긴(sparse) 측정을 획득하는 경향이 있다. 또한, 이들은 백색

벽과 같은 부드럽고 질감이 없는(textureless) 표면을 이미징하는 때에 잘 동작하지 않는다. 일부 깊이 카메라는 이미징 장치로서의 센서와 그로부터 떨어진 광 보호 장치로서의 조명기 사이의 패럴랙스(parallax)에 의해 생성된 왜곡을 측정/식별하기 위해 구조화된 광을 사용한다. 이 접근은 센서에게 보일 수 있지만 조명기에는 보이지 않는 그늘진 위치 때문에 놓친 정보로 깊이 맵을 내부적으로(inherently) 산출한다. 또한, 외부적 광은 가끔 카메라에게 보일 수 있는 구성된 패턴을 만들 수 있다.

[0012] 상술한 단점은, 3D 샘플을 3개의 깊이 카메라가 사용되는 것처럼 효과적으로 추출하기 위해 단일 조명 장치와 함께 2 이상의 센서의 배치(constellation)를 이용함으로써 극복될 수 있다. 구조광 패턴에 매칭시킴으로써 2개의 센서가 깊이 데이터를 제공할 수 있는 한편, 스테레오 기술을 적용함으로써 2개의 센서로부터의 2개의 이미지를 매칭시킴으로써 세 번째 카메라가 달성된다. 데이터 융합(fusion)을 적용함으로써, 카메라간 혼란(disruption)에 대한 강인성을 포함하여 3D 측정의 강인성(robustness)을 개선하는 것이 가능하다. 우리는 구조화된 광 기술을 이용하여, 구조화된 광 기술을 스테레오 기술과 결합하여, 그리고 감소된 오클루전과 개선된 강인성을 갖는 3D 이미지를 달성하기 위한 융합 프로세스에서 상기의 것을 이용하여, 2개의 깊이 맵을 달성하도록 단일의 프로젝터와 함께 2개 센서를 사용하는 것을 제공한다.

[0013] 도 1은 사용자의 집에서와 같이 사람(8)이 애플리케이션과 상호작용하는 모션 캡처 시스템의 예시적인 실시형태를 도시한다. 모션 캡처 시스템(10)은 디스플레이(196), 깊이 카메라 시스템(20) 및 컴퓨팅 환경 또는 장치(12)를 포함한다. 깊이 카메라 시스템(20)은 IR(적외선) 광 에미터와 같은 조명기(26), 적외선 카메라와 같은 이미지 센서(26) 및 컬러 (적-녹-청 RGB) 카메라(28)를 갖는 이미징 컴포넌트(22)를 포함할 수 있다. 사용자, 인간 또는 플레이어라고도 지칭되는 사람(8)과 같은 하나 이상의 물체는 깊이 카메라의 시야(6)에 서 있다. 라인(2와 4)은 시야(6)의 경계를 나타낸다. 이 예에서, 깊이 카메라 시스템(20) 및 컴퓨팅 환경(12)은 디스플레이(196) 상의 아바타(197)가 사람(8)의 움직임을 추적하는 애플리케이션을 제공한다. 예를 들어, 사람이 팔을 드는 때에 아바타가 팔을 들 수 있다. 아바타(197)는 3-D 가상 세계에서 길(198) 위에 서 있다. 깊이 카메라 시스템(20)의 초점 길이를 따라, 예를 들어, 수평으로 연장되는 z-축, 수직으로 연장되는 y-축 및 길이방향으로(laterally) 그리고 수평으로 연장되는 x-축을 포함하는 데카르트 세계 좌표 시스템이 정의될 수 있다. 디스플레이(196)가 y-축 방향으로 수직으로 연장되고 z-축은 y-축과 x-축에 수직으로, 그리고 사용자(8)가 서있는 지표면에 평행하게 깊이 카메라 시스템으로부터 바깥쪽으로 연장하므로, 도면의 관점은 단순화로서 수정되어 있음을 유의하라.

[0014] 일반적으로, 모션 캡처 시스템(10)은 하나 이상의 사람 타겟을 인식, 분석 및/또는 추적하는데 사용된다. 컴퓨팅 환경(12)은 애플리케이션을 실행하기 위해 컴퓨터, 게임 시스템 또는 콘솔 등뿐만 아니라 하드웨어 컴포넌트 및/또는 소프트웨어 컴포넌트를 포함할 수 있다.

[0015] 깊이 카메라 시스템(20)은, 사람에 의해 수행되는 제스처 및/또는 움직임이 캡처, 분석 및 추적되어 아바타나 스크린 상의 캐릭터를 움직이거나 사용자 인터페이스(UI)에서 메뉴 아이템을 선택하는 등 애플리케이션 내의 하나 이상의 제어나 동작을 수행할 수 있도록 사람(8)과 같은 하나 이상의 사람을 시각적으로 모니터링하는데 사용될 수 있다. 깊이 카메라 시스템(20)은 아래에서 더 자세히 논의된다.

[0016] 모션 캡처 시스템(10)은 디스플레이(196), 예를 들어, 텔레비전, 모니터, 고해상도 텔레비전(HDTV) 등, 또는 심지어 사용자에게 시각 및 청각 출력을 제공하는 벽이나 다른 표면상의 프로젝션과 같은 시청각 장치에 접속될 수 있다. 청각 출력은 별도의 장치를 통해 제공될 수도 있다. 디스플레이를 구동하기 위해, 컴퓨팅 환경(12)은 애플리케이션과 연관되는 시청각 신호를 제공하는 그래픽 카드 등의 비디오 어댑터 및/또는 사운드 카드 등의 오디오 어댑터를 포함할 수 있다. 디스플레이(196)는 컴퓨팅 환경(12)에 접속될 수 있다.

[0017] 사람(8)은 깊이 카메라 시스템(20)을 이용하여 추적될 수 있으며, 사용자의 제스처 및/또는 움직임은 아바타나 스크린 상의 캐릭터를 움직이기 위해 캡처되고 사용되며 또는 컴퓨터 환경(12)에 의해 실행되고 있는 애플리케이션에 대한 입력 제어로서 해석된다.

[0018] 사람(8)의 일부 움직임은 아바타를 제어하는 것 이외의 동작에 대응할 수 있는 제어로 해석될 수 있다. 예를 들어, 일 실시형태에서, 플레이어는 게임 종료, 중지 또는 저장, 레벨 선택, 하이스코어 보기, 친구와 통신 등을 위해 움직임을 이용할 수 있다. 플레이어는 메인 사용자 인터페이스로부터 게임 또는 기타 애플리케이션을 선택하거나 기타 옵션의 메뉴를 탐색(navigate)하기 위해 움직임을 이용할 수 있다. 그러므로, 애플리케이션과 상호작용하기 위해 적당한 여하한 방식으로 사람(8)의 움직임의 전체 범위가 사용가능하고, 사용되고, 분석될 수 있다.

- [0019] 모션 캡처 시스템(10)은 엔터테인먼트와 레저를 위한 것인 게임이나 기타 애플리케이션의 영역(realm) 밖에 있는 운영 체제 및/또는 애플리케이션 제어로 타겟 움직임을 해석하는데 더 사용될 수 있다. 예를 들어, 가상적으로 운영 체제 및/또는 애플리케이션의 여하한 제어가능한 태양이 사람(8)의 움직임에 의해 제어될 수 있다.
- [0020] 도 2는 도 1의 모션 캡처 시스템의 예시적인 블록도를 도시한다. 깊이 카메라 시스템(20)은 깊이 값을 포함할 수 있는 깊이 이미지를 포함하는 깊이 정보로써, 예를 들어, 비행시간(time-of-flight), 구조광, 스테레오 이미지 등을 포함하는 여하한 적당한 기술을 통해 비디오를 캡처하도록 구성될 수 있다. 깊이 카메라 시스템(20)은 "Z 층(Z layers)" 또는 깊이 카메라로부터 시선을 따라 연장하는 Z 축에 수직일 수 있는 층으로 깊이 정보를 조직(organize)할 수 있다.
- [0021] 깊이 카메라 시스템(20)은 물리적 공간에서 장면의 깊이 이미지를 포착하는 이미징 컴포넌트(22)를 포함할 수 있다. 깊이 이미지 또는 깊이 맵은 캡처된 장면의 2차원(2-D) 픽셀 영역을 포함할 수 있는데, 2-D 픽셀 영역 내 각각의 픽셀은 이미징 컴포넌트(22)로부터 물체까지의 직선 거리를 나타내는 관련 깊이 값을 가져서 3-D 깊이 이미지를 제공한다.
- [0022] 이미징 컴포넌트(22)의 다양한 구성이 가능하다. 한 가지 접근에서, 이미징 컴포넌트(22)는 조명기(26), 제1 이미지 센서(S1)(24), 제2 이미지 센서(S2)(29), 및 가시 컬러 카메라(28)를 포함한다. 센서(S1 및 S2)는 장면의 깊이 이미지를 캡처하는데 사용될 수 있다. 한 가지 접근에서, 조명기(26)는 적외선(IR) 광 에미터이고, 제1 및 제2 센서는 적외선 센서이다. 3-D 깊이 카메라는 조명기(26)와 하나 이상의 센서의 조합에 의해 형성된다.
- [0023] 깊이 맵은 다양한 기술을 이용하여 각각의 센서에 의해 획득될 수 있다. 예를 들어, 깊이 카메라 시스템(20)은 깊이 정보를 캡처하기 위해 구조광을 사용할 수 있다. 이러한 분석에서, 패턴닝된 광(즉, 그리드 패턴이나 스트라이프 패턴과 같은 알려진 패턴으로 표시된 광)이 조명기(26)에 의해 장면에 프로젝션된다. 장면 내의 하나 이상의 타겟 또는 물체의 표면에 부딪히는 때에, 응답으로 패턴이 변형될 수 있다. 이러한 패턴의 변형은 예를 들어, 센서(24 또는 29) 및/또는 컬러 카메라(28)에 의해 캡처될 수 있고 그 후에 깊이 카메라 시스템으로부터 타겟 또는 물체 상의 특정 위치까지의 물리적 거리를 결정하기 위해 분석될 수 있다.
- [0024] 한 가지 가능한 접근에서, 센서(24 및 29)는 조명기(26)의 반대측에, 조명기로부터 상이한 베이스라인 거리에 위치된다. 예를 들어, 센서(24)는 조명기(26)로부터 거리(BL1)에 위치되고, 센서(29)는 조명기(26)로부터 거리(BL2)에 위치된다. 센서와 조명기 사이의 거리는 센서와 조명기의 광 축과 같은 중심점 사이의 거리로 표현될 수 있다. 조명기의 반대측 상에 센서를 갖는 것의 한 가지 장점은, 센서가 상이한 관점에서 물체를 보기 때문에 시야 내 물체의 오클루전된 영역이 감소 또는 제거될 수 있다는 것이다. 또한, 센서를 조명기에 상대적으로 가깝게 배치함으로써 센서가 시야 내에서 더 가까운 물체를 보는데 최적화될 수 있는 반면, 다른 센서는 조명기로부터 상대적으로 멀리 배치함으로써 시야에서 더 먼 물체를 보는데 최적화될 수 있다. 예를 들어, BL2>BL1이면, 센서(24)는 짧은 범위 이미징에 최적화된 것으로 생각될 수 있는 반면, 센서(29)는 긴 거리 이미징에 최적화된 것으로 생각될 수 있다. 한 가지 접근에서, 센서(24와 29)는 이들이 조명기를 통과하는 공통의 선을 따라 배치되도록 동일선상에 있을 수 있다. 그러나, 센서(24와 29)의 위치에 대한 다른 구성이 가능하다.
- [0025] 예를 들어, 센서는 스캐닝될 물체 주변에, 또는 홀로그램이 프로젝션될 위치 주변에 배열될 수 있다. 또한 각각 조명기와 센서를 갖는 다수의 깊이 카메라 시스템을 물체 주변에 배열하는 것도 가능하다. 이는 물체의 상이한 측면의 보기를 허용할 수 있어서, 객체 주변의 회전 보기를 제공한다. 더 많은 깊이 카메라를 사용함으로써, 우리는 물체의 더 많은 가시 영역을 부가한다. 2개의 깊이 카메라가 그들의 조명에 의해 서로를 가리지 않는 이상, 2개의 깊이 카메라를 서로를 향하여 하나는 물체의 전면에 하나는 물체의 후면에 가질 수 있다. 각각의 깊이 카메라는 물체로부터 반사되는 그 자신의 구조광 패턴을 감지할 수 있다. 다른 예에서, 2개의 깊이 카메라가 서로에 대해 90도로 배열된다.
- [0026] 깊이 카메라 시스템(20)은 3-D 깊이 카메라(22)와 통신하는 프로세서(32)를 포함할 수 있다. 프로세서(32)는, 예를 들어, 깊이 이미지를 수신하는 것, 깊이 이미지에 기초하여 복셀(voxel)의 그리드를 생성하는 것, 사람 타겟과 연관된 하나 이상의 복셀을 고립시키기 위해 복셀의 그리드에 포함된 배경을 제거하는 것, 고립된 사람 타겟의 하나 이상의 극한의 위치 또는 포지션을 결정하는 것, 하나 이상의 극한의 위치 또는 포지션의 기초하여 모델을 조정하는 것을 위한 명령을 포함하는 명령 또는 여하한 기타 적당한 명령을 실행할 수 있는 표준화된 프로세서, 특화된 프로세서, 마이크로프로세서 등을 포함할 수 있으며, 이는 아래에서 더 상세히 설명할 것이다.

- [0027] 프로세서(32)는 구조광 깊이 맵을 도출하는 소프트웨어(33), 스테레오스코픽 비전 깊이 맵을 도출하는 소프트웨어(34), 깊이 맵 병합 계산을 수행하는 소프트웨어(35)를 사용하기 위해 메모리(31)에 액세스할 수 있다. 프로세서(32)는 조명 평면에서 조명기에 의해 발산되는 구조광의 패턴에 픽셀 데이터의 프레임에 비교함으로써 객체의 구조된 광 깊이 맵을 도출하는 적어도 하나의 제어 회로인 것으로 생각될 수 있다. 예를 들어, 소프트웨어(33)를 이용하여, 적어도 하나의 제어 회로는, 센서(24)에 의해 획득되는 픽셀 데이터의 제1 프레임을 조명기(26)에 의해 발산되는 구조광의 패턴에 비교함으로써 물체의 제1 구조광 깊이 맵을 도출할 수 있고, 센서(29)에 의해 획득되는 픽셀 데이터의 제2 프레임을 구조광의 패턴에 비교함으로써 물체의 제2 구조광 깊이 맵을 도출할 수 있다. 적어도 하나의 제어 회로는 제1 및 제2 구조광 깊이 맵에 기초하는 병합된 깊이 맵을 도출하기 위해 소프트웨어(35)를 사용할 수 있다. 구조광 깊이 맵은, 예를 들어, 도 5(a)와 관련하여 아래에서 더 논의된다.
- [0028] 또한, 적어도 하나의 제어 회로는, 센서(24)에 의해 획득된 픽셀 데이터의 제1 프레임을 센서(29)에 의해 획득된 픽셀 데이터의 제2 프레임에 스테레오스코픽 매칭함으로써 적어도 물체의 제1 스테레오스코픽 깊이 맵을 도출하고, 픽셀 데이터의 제2 프레임을 픽셀 데이터의 제1 프레임에 스테레오스코픽 매칭함으로써 적어도 물체의 제2 스테레오스코픽 깊이 맵을 도출하기 위해 소프트웨어(34)를 이용할 수 있다. 소프트웨어(25)는 하나 이상의 구조광 깊이 맵 및/또는 스테레오스코픽 깊이 맵을 병합할 수 있다. 스테레오스코픽 깊이 맵은 예를 들어 도 5(b)와 관련하여 아래에서 더 논의된다.
- [0029] 적어도 하나의 제어 회로가 프로세서(192) 또는 여하한 기타 프로세서와 같은 깊이 카메라 시스템 외부의 프로세서에 의해서도 제공될 수 있다. 적어도 하나의 제어 회로는 메모리(31)로부터의 소프트웨어에 액세스할 수 있는데, 메모리(31)는, 예를 들어, 적어도 하나의 프로세서나 제어기(32)를 여기에 설명된 깊이 카메라 시스템에서 이미지 데이터를 처리하는 방법을 수행하도록 프로그래밍하기 위한 컴퓨터 판독가능 소프트웨어가 포함된 유형의(tangible) 컴퓨터 판독가능 저장소일 수 있다.
- [0030] 메모리(31)는 프로세서(32)에 의해 실행되는 명령을 저장할 수 있을 뿐만 아니라, 센서나 컬러 카메라에 의해 캡처된 픽셀 데이터(36)의 프레임과 같은 이미지를 저장할 수 있다. 예를 들어, 메모리(31)는 RAM(random access memory), ROM(read only memory), 캐시, 플래시 메모리, 하드 디스크, 또는 여하한 기타의 적합한 유형의(tangible) 컴퓨터 판독가능 저장 컴포넌트를 포함할 수 있다. 메모리 컴포넌트(31)는 버스(21)를 통해 이미지 캡처 컴포넌트(22) 및 프로세서(32)와 통신하는 별도의 컴포넌트일 수 있다. 다른 실시형태에 따르면, 메모리 컴포넌트(31)는 프로세서(32) 및/또는 이미지 캡처 컴포넌트(22)로 통합될 수 있다.
- [0031] 깊이 카메라 시스템(20)은 유선 및/또는 무선 접속과 같은 통신 링크(37)를 통해 컴퓨팅 환경(12)과 통신할 수 있다. 컴퓨팅 환경(12)은 깊이 카메라 시스템(20)의 시야에 있는 물리적 공간으로부터 언제 이미지 데이터를 캡처할지를 나타내는 클럭 신호를 통신 링크(37)를 통해 깊이 카메라 시스템(20)에 제공할 수 있다.
- [0032] 또한, 깊이 카메라 시스템(20)은, 예를 들어, 이미지 센서(24와 29) 및/또는 컬러 카메라(28)에 의해 캡처되는 이미지와 깊이 정보, 및/또는 깊이 카메라 시스템(20)에 의해 생성될 수 있는 골격 모델(skeletal model)을 통신 링크(37)를 통해 컴퓨팅 환경(12)에 제공할 수 있다. 그러면 컴퓨팅 환경(12)은 애플리케이션을 제어하기 위해 모델, 깊이 정보 및 캡처된 이미지를 이용할 수 있다. 예를 들어, 도 2에 도시된 바와 같이, 컴퓨팅 환경(12)은, 각각이 (사용자가 움직임에 따라) 골격 모델에 의해 수행될 수 있는 제스처에 관한 정보를 갖는 제스처 필터의 집합과 같은 제스처 라이브러리(190)를 포함할 수 있다. 예를 들어, 손으로 휘두르거나(swiping) 던지는 것(flinging)과 같은 다양한 손 제스처에 대해 제스처 필터가 제공될 수 있다. 검출된 동작을 각각의 필터에 비교함으로써, 사람에 의해 수행되는 구체적인 제스처나 움직임이 식별될 수 있다. 움직임이 수행되는 범위로 결정될 수 있다.
- [0033] 깊이 카메라 시스템(20)에 의해 캡처되는 골격 모델 형태의 데이터와 그와 연관된 움직임은, (골격 모델로 나타내어지는) 사용자가 하나 이상의 구체적인 움직임을 언제 수행했는지를 식별하기 위해 제스처 라이브러리(190) 내의 제스처 필터들에 비교될 수 있다. 이들 움직임은 애플리케이션의 다양한 제어와 연관될 수 있다.
- [0034] 컴퓨팅 환경은 디스플레이 장치(196)에 시청각 출력 신호를 제공하고 여기에 설명된 다른 기능을 달성하기 위해 메모리(194)에 저장되는 명령을 실행하기 위한 프로세서(192)도 포함할 수 있다.
- [0035] 도 3은 도 1의 모션 캡처 시스템에서 사용될 수 있는 컴퓨팅 환경의 예시적인 블록도를 도시한다. 컴퓨팅 환경은 하나 이상의 제스처나 기타 움직임을 해석하고, 그에 응답하여 디스플레이 상에서 시각적 공간을 업데이트하는데 사용될 수 있다. 상술한 컴퓨팅 환경(12)과 같은 컴퓨팅 환경은 게임 콘솔과 같은 멀티미디어 콘솔(100)

을 포함할 수 있다. 멀티미디어 콘솔(100)은 레벨 1 캐시(102), 레벨 2 캐시(104) 및 플래시 ROM(Read Only Memory)(106)를 갖는 CPU(central processing unit)(101)를 갖는다. 레벨 1 캐시(102)와 레벨 2 캐시(104)는 데이터를 일시적으로 저장하고, 따라서 메모리 액세스 사이클의 수를 감소시키며, 그에 의해 처리 스피드와 스루풋(throughput)을 개선한다. 하나 이상의 코어를 갖고, 따라서, 추가의 레벨 1 및 레벨 2 캐시(102와 104)를 갖는 CPU(101)가 제공될 수 있다. 플래시 ROM과 같은 메모리(106)는 멀티미디어 콘솔(100)이 켜지는 때에 부팅 프로세스의 초기 페이지 동안 로딩되는 실행 코드를 저장할 수 있다.

[0036] GPU(graphic processing unit)(그래픽 처리 유닛)(108) 및 비디오 인코더/비디오 코덱(코더/디코더)(114)은 고속 및 고해상도 그래픽 처리를 위한 비디오 처리 파이프라인을 형성한다. 데이터는 버스를 통해 그래픽 처리 유닛(108)으로부터 비디오 인코더/비디오 코덱(114)으로 운반된다. 비디오 처리 파이프라인은 텔레비전 또는 기타 디스플레이로의 전송을 위해 데이터를 A/V(오디오/비디오) 포트(140)로 출력한다. 프로세서가 RAM(Random Access Memory)과 같은 다양한 타입의 메모리(112)에 액세스하는 것을 용이하게 하도록 메모리 제어기(110)가 GPU(108)에 접속된다.

[0037] 멀티미디어 콘솔(100)은 I/O 제어기(120), 시스템 관리 제어기(122), 오디오 처리 유닛(123), 네트워크 인터페이스(124), 제1 USB 호스트 제어기(126), 제2 USB 제어기(128), 및 모듈(118) 상에서 구현되는 것이 바람직한 프론트 패널 I/O 서브어셈블리(130)를 포함한다. USB 제어기(126과 128)는 주변 제어기(142(1)-142(2)), 무선 어댑터(148) 및 외부 메모리 장치(146)(예를 들어, 플래시 메모리, 외부 CD/DVD ROM 드라이브, 제거가능(removable) 매체 등)에 대한 호스트로서 기능한다. 네트워크 인터페이스(NW IF)(124) 및/또는 무선 어댑터(148)는 네트워크(예를 들어, 인터넷, 홈네트워크 등)으로의 액세스를 제공하고, 이더넷 카드, 모뎀, 블루투스 모듈, 케이블 모뎀 등을 포함하는 다양한 유선이나 무선 어댑터 컴포넌트 중 여하한 것일 수 있다.

[0038] 부팅 프로세스 동안 로딩되는 애플리케이션 데이터를 저장하기 위해 시스템 메모리(143)가 제공된다. 미디어 드라이브(144)가 제공되며, DVD/CD 드라이브, 하드 드라이브 또는 기타 제거가능 미디어 드라이브를 포함할 수 있다. 미디어 드라이브(144)는 멀티미디어 콘솔(100)의 내부 또는 외부에 있을 수 있다. 애플리케이션 데이터는 멀티미디어 콘솔(100)에 의한 실행, 플레이백 등을 위하여 미디어 드라이브(144)를 통해 액세스될 수 있다. 미디어 드라이브(144)는 직렬 ATA 버스 또는 기타 고속 접속과 같은 버스를 통해 I/O 제어기(120)에 접속된다.

[0039] 시스템 관리 제어기(122)는 멀티미디어 콘솔(100)의 가용성을 보장하는 것과 관련된 다양한 서비스 기능을 제공한다. 오디오 처리 유닛(123)과 오디오 코덱(132)은 고충실(high fidelity) 및 스테레오 처리를 하는 대응하는 오디오 처리 파이프라인을 형성한다. 오디오 데이터는 통신 링크를 통해 오디오 처리 유닛(123)과 오디오 코덱(132) 사이에서 운반된다. 오디오 처리 파이프라인은 오디오 기능을 갖는 외부 오디오 플레이어나 장치에 의한 재생을 위해 데이터를 A/V 포트(140)로 출력한다.

[0040] 프론트 패널 I/O 서브어셈블리(130)는 전원 버튼(150)과 꺼냄(eject) 버튼(152) 뿐만 아니라 여하한 LED(light emitting diodes) 또는 멀티미디어 콘솔(100)의 외부 표면에 노출된 기타 표시자의 기능을 지원한다. 시스템 전원 모듈(136)은 멀티미디어 콘솔(100)의 컴포넌트에 전원을 제공한다. 팬(138)은 멀티미디어 콘솔(100) 내의 회로를 냉각한다.

[0041] CPU(101), GPU(108), 메모리 제어기(110) 및 멀티미디어 콘솔(100) 내 다양한 기타 컴포넌트는, 직렬 및 병렬 버스, 메모리 버스, 주변 버스 및 다양한 버스 아키텍처 중 여하한 것을 이용하는 프로세서 또는 로컬 버스를 포함하는 하나 이상의 버스를 통해 상호접속된다.

[0042] 멀티미디어 콘솔(100)이 켜지면, 애플리케이션 데이터가 시스템 메모리(143)로부터 메모리(112) 및/또는 캐시(102, 104)로 로딩되고 CPU(101) 상에서 실행될 수 있다. 애플리케이션은 멀티미디어 콘솔(100) 상에서 사용가능한 상이한 미디어 타입으로 탐색하는 때에 일관적인 사용자 경험을 제공하는 그래픽 사용자 인터페이스를 제시할 수 있다. 동작에서, 애플리케이션 및/또는 미디어 드라이브(144) 내에 포함된 기타 매체는 멀티미디어 콘솔(100)에 추가적인 기능을 제공하기 위해 미디어 드라이브(144)로부터 론칭 또는 플레이될 수 있다.

[0043] 멀티미디어 콘솔(100)은 시스템을 텔레비전이나 기타 디스플레이에 단순히 접속함으로써 단독(standalone) 시스템으로 동작될 수 있다. 이 단독 모드에서, 멀티미디어 콘솔(100)은 하나 이상의 사용자가 시스템과 상호작용하거나, 영화를 보거나, 음악을 들을 수 있도록 한다. 그러나, 네트워크 인터페이스(124) 또는 무선 어댑터(148)를 통해 사용가능하게 되는 브로드밴드 접속의 통합에 의해, 멀티미디어 콘솔(100)은 또한 더 큰 네트워크 커뮤니티의 참가자로서 동작될 수 있다.

- [0044] 멀티미디어 콘솔(100)이 켜지는 때에, 하드웨어 리소스의 특정량이 멀티미디어 콘솔 운영 체제에 의한 시스템 사용을 위해 유보된다. 이들 리소스는 메모리(예를 들어, 16MB), CPU와 GPU 사이클(예를 들어, 5%), 네트워킹 대역폭(예를 들어, 8kb) 등의 유보를 포함할 수 있다. 이들 리소스가 시스템 부팅 시간에 유보되므로, 유보된 리소스는 애플리케이션의 관점에서는 존재하지 않는다.
- [0045] 구체적으로, 메모리 유보는 론칭 커널, 동시적 시스템 애플리케이션과 드라이버를 포함하기에 충분하게 큰 것이 바람직하다. 유보된 CPU 사용이 시스템 애플리케이션에 의해 사용되지 않는 경우 유휴 스레드(thread)가 여하한 미사용 사이클을 소비하도록, CPU 유보는 바람직하게는 일정하다.
- [0046] GPU 유보와 관련하여, 팝업을 오버레이로 렌더링하기 위한 코드를 스케줄하도록 GPU 인터럽트를 이용함으로써 시스템 애플리케이션에 의해 생성되는 가벼운 메시지(예를 들어, 팝업)가 표시된다. 오버레이에 대해 요구되는 메모리의 양은 오버레이 영역 크기에 의존하고, 오버레이는 바람직하게는 스크린 해상도에 의해 스케일링(scale)된다. 완전한 사용자 인터페이스가 동시적 시스템 애플리케이션에 의해 사용되는 경우, 애플리케이션 해상도에 독립적인 해상도를 사용하는 것이 바람직하다. 스케일링은 주파수를 변경하고 TV 재동기화(resynch)를 유발할 필요가 없도록 이 해상도를 설정하는데 사용될 수 있다.
- [0047] 멀티미디어 콘솔(100)이 부팅되고 시스템 리소스가 유보된 후에, 동시적 시스템 애플리케이션은 시스템 기능을 제공하도록 실행된다. 시스템 기능은 상술한 유보된 시스템 리소스 내에서 실행되는 시스템 애플리케이션의 세트에 캡슐화된다. 운영 체제 커널은, 게임 애플리케이션 스레드에 대해 시스템 애플리케이션 스레드인 스레드를 식별한다. 애플리케이션에 일관적인 시스템 리소스 모습을 제공하기 위해 시스템 애플리케이션은 바람직하게는 사전결정된 시간에 사전결정된 간격으로 CPU(101) 상에서 실행되도록 스케줄된다. 스케줄링은 콘솔 상에서 실행되는 게임 애플리케이션에 대한 캐시 혼란을 최소화할 것이다.
- [0048] 동시적 시스템 애플리케이션이 오디오를 요구하는 때에, 시간 민감성 때문에 오디오 처리는 게임 애플리케이션과 비동기적으로 스케줄된다. 멀티미디어 콘솔 애플리케이션 관리자(후술함)는 시스템 애플리케이션이 활성인 때에 게임 애플리케이션 오디오 수준(예를 들어, 무음, 감쇠(attenuate))을 제어한다.
- [0049] 입력 장치(예를 들어, 제어기(142(1)과 142(2)))는 게임 애플리케이션과 시스템 애플리케이션에 의해 공유된다. 입력 장치는 유보된 리소스가 아니고, 시스템 애플리케이션과 게임 애플리케이션 사이에서 각각이 장치의 포커스를 갖도록 전환될 것이다. 애플리케이션 관리자는 바람직하게는 게임 애플리케이션의 지식을 알지 못하고 입력 스트림의 전환을 제어하고 드라이버는 포커스 전환에 대한 상태 정보를 유지한다. 콘솔(100)은 센서(24와 29)를 포함하는 도 2의 깊이 카메라 시스템으로부터 추가 입력을 수신한다.
- [0050] 도 4는 도 1의 모션 캡처 시스템에서 사용될 수 있는 컴퓨팅 환경의 다른 예시적인 블록도를 도시한다. 모션 캡처 시스템에서, 컴퓨팅 환경은 하나 이상의 제스처 또는 기타 움직임을 해석하고, 그에 응답하여 디스플레이 상의 시각적 공간을 업데이트하는데 사용될 수 있다. 컴퓨팅 환경(220)은 컴퓨터(241)를 포함하는데, 컴퓨터는 통상 다양한 유형의(tangible) 컴퓨터 판독가능 저장 매체를 포함한다. 이는 컴퓨터(214)에 의해 액세스될 수 있고 휘발성 및 비휘발성 매체, 제거가능 및 비제거가능 매체 모두를 포함하는 여하한 가용 매체일 수 있다. 시스템 메모리(222)는 ROM(read only memory)(223)과 RAM(random access memory)(260)과 같은 휘발성 및/또는 비휘발성 메모리의 형태로 컴퓨터 저장 매체를 포함한다. 스타트업 동안 등에 컴퓨터(241) 내의 요소들 사이에서 정보를 전달하는 것을 돕는 기본 루틴을 포함하는 기본 입력/출력 시스템(224)(BIOS)은 통상 ROM(223)에 저장된다. RAM(260)은 통상, 처리 유닛(259)에 의해 즉시 액세스가능하고 및/또는 현재 동작되고 있는 데이터 및/또는 프로그램 모듈을 포함한다. 그래픽 인터페이스(231)는 GPU(229)와 통신한다. 제한이 아니라 예로써, 도 4는 운영 체제(225), 애플리케이션 프로그램(226), 기타 프로그램 모듈(227) 및 프로그램 데이터(228)을 도시한다.
- [0051] 컴퓨터(241)는 기타 제거가능/비제거가능, 휘발성/비휘발성 컴퓨터 저장 매체, 예를 들어, 비제거가능 비휘발성 자기 매체로부터 판독하거나 이에 기록하는 하드 디스크 드라이브(238), 제거가능 비휘발성 자기 디스크(254)로부터 판독하거나 이에 기록하는 자기 디스크 드라이브(239), 및 CD ROM이나 기타 광 매체와 같은 제거가능 비휘발성 광 디스크(253)로부터 판독하거나 이에 기록하는 광 디스크 드라이브(240)도 포함할 수 있다. 예시적인 동작 환경에서 사용될 수 있는 기타의 제거가능/비제거가능, 휘발성/비휘발성 유형의(tangible) 컴퓨터 판독가능 저장 매체는 자기 테이프 카세트, 플래시 메모리 카드, DVD(digital versatile disk), 디지털 비디오 테이프, 고체 상태(solid state) RAM, 고체 상태 ROM 등을 포함하지만 이에 제한되지 않는다. 하드 디스크 드라이브(238)는 통상, 인터페이스(234)와 같은 비제거가능 메모리 인터페이스를 통해 시스템 버스(221)에 접속되고, 자기 디스크 드라이브(239)와 광 디스크 드라이브(240)는 통상, 인터페이스(235)와 같은 제거가능 메모리 인터페

이스에 의해 시스템 버스(221)에 접속된다.

[0052] 위에서 설명하고 도 4에 도시된 드라이브와 그들의 관련 컴퓨터 저장 매체는 컴퓨터 판독가능 명령, 데이터 구조, 프로그램 모듈 및 컴퓨터(241)를 위한 기타 데이터의 저장을 제공한다. 예를 들어, 하드 디스크 드라이브(238)는 운영 체제(258), 애플리케이션 프로그램(257), 기타 프로그램 모듈(256) 및 프로그램 데이터(255)를 저장하는 것으로 도시된다. 이들 컴포넌트는 운영 체제(225), 애플리케이션 프로그램(226), 기타 프로그램 모듈(227) 및 프로그램 데이터(228)와 동일하거나 다를 수 있음을 유의하라. 운영 체제(258), 애플리케이션 프로그램(257), 기타 프로그램 모듈(256) 및 프로그램 데이터(255)에는, 최소한 그들이 상이한 사본임을 도시하기 위해 여기서 상이한 번호가 주어진다. 사용자는 키보드(251) 및 보통 마우스, 트랙볼 또는 터치 패드라고 불리는 포인팅 장치(252)와 같은 입력 장치를 통해 컴퓨터(241)로 명령과 정보를 입력할 수 있다. 다른 입력 장치(미도시)는 마이크론, 조이스틱, 게임 패드, 위성 접시, 스캐너 등을 포함할 수 있다. 이들 및 다른 입력 장치가 종종 시스템 버스에 결합되는 사용자 입력 인터페이스(236)를 통해 처리 유닛(259)에 접속되지만, 병렬 포트, 게임 포트 또는 USB(universal serial bus)와 같은 다른 인터페이스와 버스 구조에 의해 접속될 수도 있다. 센서(24와 29)를 포함하는 도 2의 깊이 카메라 시스템(20)은 콘솔(100)에 대한 추가의 입력 장치를 정의할 수 있다. 모니터(242) 또는 다른 유형의 디스플레이도 비디오 인터페이스(232)와 같은 인터페이스를 통해 시스템 버스(221)에 접속된다. 모니터에 추가하여, 컴퓨터는 스피커(244)와 프린터(243)와 같은 다른 주변 출력 장치도 포함할 수 있고, 이는 출력 주변 인터페이스(233)를 통해 접속될 수 있다.

[0053] 컴퓨터(241)는 원격 컴퓨터(246)와 같은 하나 이상의 원격 컴퓨터로의 논리적 접속을 이용하여 네트워크된 환경에서 동작할 수 있다. 원격 컴퓨터(246)는 퍼스널 컴퓨터, 서버, 라우터, 네트워크 PC, 피어 장치 또는 기타 공통 네트워크 노드일 수 있고, 보통, 컴퓨터(241)에 관하여 상술한 요소 중 많은 것 또는 전부를 포함하지만, 도 4에는 메모리 저장 장치(247)만이 도시되었다. 논리적 접속은 LAN(local area network)(245)과 WAN(wide area network)(249)를 포함하지만, 다른 네트워크도 포함할 수 있다. 이러한 네트워킹 환경은 사무실, 기업 범위 컴퓨터 네트워크, 인트라넷 및 인터넷에서 흔하다.

[0054] LAN 네트워킹 환경에서 사용되는 때에, 컴퓨터(241)는 네트워크 인터페이스 또는 어댑터(237)를 통해 LAN(245)에 접속된다. WAN 네트워킹 환경에서 사용되는 때에, 컴퓨터(241)는 보통, 모뎀(250) 또는 인터넷과 같은 WAN(249)을 통해 통신을 수립하기 위한 기타 수단을 포함한다. 내부 또는 외부에 있을 수 있는 모뎀(250)은 사용자 인터페이스(236) 또는 기타 적당한 매커니즘을 통해 시스템 버스(221)에 접속될 수 있다. 네트워킹된 환경에서, 컴퓨터(241)에 대해 도시된 프로그램 모듈 또는 그 부분은 원격 메모리 저장 장치에 저장될 수 있다. 제한이 아니라 예로서, 도 4는 원격 애플리케이션 프로그램(248)을 메모리 장치(247) 상에 상주하는 것으로 도시한다. 도시된 네트워크 접속은 예시적인 것이고 컴퓨터 사이에 통신 링크를 수립하는 다른 수단이 사용될 수 있음을 인식할 것이다.

[0055] 컴퓨팅 환경은 여기서 설명된 깊이 카메라 시스템에서 이미지 데이터를 처리하기 위한 방법을 수행하도록 적어도 하나의 프로세서를 프로그래밍하기 위한 컴퓨터 판독가능 소프트웨어를 포함하는 유형의(tangible) 컴퓨터 판독가능 저장소를 포함할 수 있다. 유형의 컴퓨터 판독가능 저장소는, 예를 들어, 컴포넌트(31, 194, 222, 234, 235, 230, 253 및 254) 중 하나 이상을 포함할 수 있다. 프로세서는, 예를 들어, 하나 이상의 컴포넌트(32, 192, 229 및 259)를 포함할 수 있다.

[0056] 도 5(a)는 구조광 시스템에서 캡처된 프레임과 조명 프레임을 도시한다. 조명 프레임(500)은 조명기의 이미지 평면을 나타내는데, 조명기는 조명기의 시야에서 물체(520)에 구조광을 발산한다. 조명 프레임(500)은 x_2 , y_2 및 z_2 수직 축을 갖는 좌표계를 포함한다. F_2 는 조명기의 초점이고, O_2 는 조명 프레임(500)의 중심에서와 같은 좌표계의 원점이다. 발산된 구조광은 스트라이프, 스폿(spot) 또는 기타 알려진 조명 패턴을 포함할 수 있다. 유사하게, 캡처된 프레임(510)은 도 2와 관련하여 논의된 센서(24 또는 29)와 같은 센서의 이미지 평면을 나타낸다. 캡처된 프레임(510)은 x_1 , y_1 및 z_1 수직 축을 갖는 좌표계를 포함한다. F_1 는 센서의 초점이고 O_1 는 캡처된 프레임(510)의 중심에서와 같은 좌표계의 원점이다. 이 예에서, 단순성을 위하여, 이것이 필요한 것은 아니지만, y_1 과 y_2 는 동일선상에 정렬되고 z_1 과 z_2 는 병렬이다. 또한, 2 이상의 센서가 사용될 수 있지만 단순화를 위해 하나의 센서만 여기 도시된다.

[0057] 조명 프레임(500) 상의 점 P_2 로부터 발산되는 예시적인 광선(502)과 같은 프로젝션된 구조광선이 조명기 평면에서 상이한 x_2 , y_2 위치로부터 발산된다. 광선(502)은 점 P_0 에서 물체(520), 예를 들어 사람에 부딪히고, 많은 방향으로 반사된다. 광선(512)은 P_0 에서 캡처된 프레임(510) 상의 점 P_1 으로 이동하는 예시적인 반사 광선이다.

P_1 은 센서 내 픽셀에 의해 나타내어져서, 그 x_1 , y_1 위치가 알려진다. 기하학 원리에 의해, P_2 는 P_1 , F_1 및 F_2 를 포함하는 평면에 있다. 조명 프레임(500)과 교차하는 이 평면의 부분은 에피-폴라(epi-polar) 라인(505)이다. 구조광의 어떤 부분이 P_2 에 의해 프로젝션되는지 식별함으로써, 에피-폴라 라인(505)을 따르는 P_2 의 위치가 식별될 수 있다. P_2 는 P_1 의 대응 점이다. 물체의 깊이가 가까울수록, 에피-폴라 라인의 길이가 길다.

[0058] 후속하여, z_1 축에 따른 P_0 의 깊이는 삼각법(triangulation)으로 결정될 수 있다. 이는 깊이 맵에서 픽셀 P_1 에 할당되는 깊이 값이다. 조명 프레임(500)의 일부 점에 대해, 오클루전에 기인하거나 센서의 제한된 시야에 기인하는 등으로 캡처된 프레임(510) 내에 대응 픽셀이 없을 수 있다. 조명 프레임(500)에서 대응 픽셀이 식별되는 캡처된 프레임(510) 내의 각각의 픽셀에 대해, 깊이 값이 획득될 수 있다. 캡처된 프레임(510)에 대한 깊이 값의 세트는 캡처된 프레임(510)에 대한 깊이 맵을 제공한다. 추가의 센서 및 그들 각각의 캡처된 프레임에 대해 유사한 프로세스가 수행될 수 있다. 또한, 비디오 데이터의 연속적인 프레임이 획득되는 때에, 각각의 프레임에 대해 프로세스가 수행될 수 있다.

[0059] 도 5(b)는 스테레오스코픽 광 시스템 내의 2개의 캡처된 프레임을 도시한다. 스테레오스코픽 처리는, 2개의 프레임에서 대응되는 점이 식별된다는 점에서 도 5(a)에서 설명된 처리와 유사하다. 그러나, 이 경우, 2개의 캡처된 프레임 내의 대응되는 픽셀이 식별되고, 조명이 별도로 제공된다. 조명기(550)는 조명기의 시야에서 물체(520)에 프로젝션된 광을 제공한다. 이 광은 예를 들어 물체에 의해 반사되고 2개의 센서에 의해 감지된다. 제1 센서는 픽셀 데이터의 프레임(530)을 획득하는 반면, 제2 센서는 픽셀 데이터의 프레임(540)을 획득한다. 예시적인 광선(532)은 물체 상의 점 P_0 에서 프레임(530) 내 픽셀 P_2 까지 연장하여, 연관된 센서의 초점 F_2 를 통과한다. 유사하게, 예시적인 광선(542)은 물체 상의 점 P_0 에서 프레임(540) 내 픽셀 P_1 까지 연장하여, 연관된 센서의 초점 F_1 를 통과한다. 프레임(540)의 관점에서, 스테레오 매칭에는 P_1 에 대응하는 에피-폴라 라인(545) 상의 점 P_2 를 식별하는 것이 관여될 수 있다. 유사하게, 프레임(530)의 관점에서, 스테레오 매칭에는 P_2 에 대응하는 에피-폴라 라인(548) 상의 점 P_1 을 식별하는 것이 관여될 수 있다. 그러므로, 스테레오 매칭은 한 쌍의 프레임 중 각 프레임에 대해 한 번씩 별도로 수행될 수 있다. 일부 경우에, 제1 프레임으로부터 제2 프레임으로의 한 방향의 스테레오 매칭이 제2 프레임으로부터 제1 프레임으로의 다른 방향의 스테레오 매칭을 수행하지 않고 수행될 수 있다.

[0060] z_1 축에 따른 P_0 의 깊이는 삼각법에 의해 결정될 수 있다. 이는 픽셀 깊이 맵에서 픽셀 P_1 에 할당되는 깊이 값이다. 프레임(540) 내 일부 점에 대해, 오클루전에 기인하거나 센서의 제한된 시야에 기인하는 등으로 프레임(530) 내에 대응 픽셀이 없을 수 있다. 프레임(530)에서 대응 픽셀이 식별되는 프레임(540) 내의 각각의 픽셀에 대해, 깊이 값이 획득될 수 있다. 프레임(540)에 대한 깊이 값의 세트는 프레임(540)에 대한 깊이 맵을 제공한다.

[0061] 유사하게, z_2 축에 따른 P_2 의 깊이는 삼각법에 의해 결정될 수 있다. 이는 픽셀 깊이 맵에서 픽셀 P_2 에 할당되는 깊이 값이다. 프레임(530) 내 일부 점에 대해, 오클루전에 기인하거나 센서의 제한된 시야에 기인하는 등으로 프레임(540) 내에 대응 픽셀이 없을 수 있다. 프레임(540)에서 대응 픽셀이 식별되는 프레임(530) 내의 각각의 픽셀에 대해, 깊이 값이 획득될 수 있다. 프레임(530)에 대한 깊이 값의 세트는 프레임(530)에 대한 깊이 맵을 제공한다.

[0062] 유사한 프로세스가 추가의 센서 및 그들 각각의 캡처된 프레임에 대해 수행될 수 있다. 또한, 비디오 데이터의 연속적인 프레임이 획득되는 때에, 각각의 프레임에 대해 프로세스가 수행될 수 있다.

[0063] 도 6(a)는 조명기의 공통 축에 2개의 센서를 갖는 이미징 컴포넌트(600)를 도시한다. 조명기(26)는 구조광 패턴으로 시야 내의 사람 타겟이나 기타 물체를 조명하는 프로젝터이다. 광원은, 예를 들어, $0.75 \mu\text{m} - 1.4 \mu\text{m}$ 의 파장을 갖는 근적외선 광, $3 \mu\text{m} - 8 \mu\text{m}$ 의 파장을 갖는 중파장 적외선 광 및 사람에 의해 발산되는 적외선 방사에 가장 가까운 열 이미징 영역인 $8 \mu\text{m} - 15 \mu\text{m}$ 의 파장을 갖는 장파장 적외선 광을 포함하는 $700 \text{ nm} - 3,000 \text{ nm}$ 의 파장을 갖는 적외선 레이저일 수 있다. 조명기는 레이저광을 수신하고 다수의 산란된 광 빔을 출력하는 DOE(diffractive optical element)를 포함할 수 있다. 일반적으로, DOE는 단일의 시준된(collimated) 광 빔으로부터의, 1000개의 작은 광 빔과 같은 다수의 작은 광 빔을 제공하는데 사용된다. 각각의 작은 광 빔은 단일의 시준된 광 빔의 전력의 작은 부분을 갖고, 작은 산란 광 빔은 명목상(nominally) 동일한 강도를 가질 수 있다.

- [0064] 작은 광 빔은 희망의 사전결정된 패턴으로 조명기의 시야를 정의한다. DOE는 빔 복제기(replicator)이고, 그래서 모든 출력 빔은 입력 빔과 동일한 기하학적 구조(geometry)를 갖는다. 예를 들어, 모션 트래킹 시스템에서, 방 안에 서거나 앉아 있는 사람 타겟의 추적을 허용하는 방식으로 방을 조명하는 것이 바람직할 수 있다. 전체 사람 타겟을 추적하기 위해, 시야는 충분히 넓은 각도와, 높이 및 폭으로 확장되어 사람의 전체 높이와 넓이 그리고 사람이 모션 트래킹 시스템의 애플리케이션과 상호작용하는 때에 움직일 수 있는 영역을 조명하여야 한다. 팔을 머리 위로 올리거나 양 쪽으로 뻗을 때의 팔 범위, 사람이 애플리케이션과 상호작용하는 때에 움직일 수 있는 영역의 크기, 카메라로부터의 사람의 기대 거리와 카메라의 초점 길이를 포함하여, 사람의 기대 키와 넓이와 같은 인자에 기초하여 적당한 시야가 설정될 수 있다.
- [0065] 진술한 RGB 카메라(28)도 제공될 수 있다. RGB 카메라는 도 6(b) 및 6c에서도 제공될 수 있지만 단순화를 위해 도시되지 않는다.
- [0066] 이 예에서, 센서(24와 29)는 조명기(26)의 공통측에 있다. 센서(24)는 조명기(26)로부터 베이스라인 거리(BL1)에 있고, 센서(29)는 조명기(26)로부터 베이스라인 거리(BL2)에 있다. 센서(29)는 더 작은 베이스라인에 의해 짧은 범위 이미징에 최적화되는 반면, 센서(24)는 더 긴 베이스라인에 의해 긴 범위 이미징에 최적화된다. 또한, 조명기의 일 측에 모든 센서를 배치함으로써, 크기가 제한된 하우스징을 보통 포함하는 이미징 컴포넌트(600)의 고정된 크기에 대해, 조명기로부터 더 먼 센서에 대해 더 긴 베이스라인이 달성될 수 있다. 반면, 더 짧은 베이스라인은, 주어진 초점 길이를 가정하면, 센서가 더 가까운 물체에 초점을 맞출 수 있어서 짧은 거리에 대해 더 정확한 깊이 측정이 가능하게 되므로, 짧은 범위 이미징을 개선한다. 더 짧은 베이스라인은 더 작은 디스패리티(disparity)와 최소 오클루전을 가져온다.
- [0067] 더 긴 베이스라인은, 대응하는 점의 광선 사이의 각이 더 크므로, 즉 이미지 픽셀이 더 작은 거리 차이를 검출할 수 있으므로, 긴 범위 이미징의 정확도를 개선한다. 예를 들어, 도 5(a)에서, 광(502와 512) 사이의 각도는 프레임(500과 510)이 더 멀리 떨어지면 더 커질 것이라는 것을 알 수 있다. 그리고, 도 5(b)에서, 광(532와 542) 사이의 각도는 프레임(530과 540)이 더 멀리 떨어지면 더 커질 것이라는 것을 알 수 있다. 센서가 더 멀리 떨어져서 광선 사이의 각도가 크면 깊이를 결정하기 위한 삼각법 프로세스가 더 정확하다.
- [0068] 짧은 범위 또는 긴 범위 이미징 중 어떤 것이 최적화되는지에 따라 센서에 대해 최적 베이스라인을 설정하는 것에 추가하여, 이미징 컴포넌트(600)의 하우스징의 제한 내에서, 센서의 다른 특성이 짧거나 긴 범위 이미징을 최적화하기 위하여 설정될 수 있다. 예를 들어, 카메라의 공간 해상도가 최적화될 수 있다. CCD(charge-couple device)와 같은 센서의 공간 해상도는 픽셀의 수와 프로젝션된 이미지에 대한 그 크기의 함수이고, 센서에 의해 세부사항이 얼마나 정밀하게 측정되는지의 척도이다. 짧은 거리 이미징에 최적화된 센서에 대해, 긴 거리 이미징에 최적화된 센서에 비해 낮은 공간 해상도가 수용가능할 수 있다. 낮은 공간 해상도는 프레임 내에서 상대적으로 적은 수의 픽셀 및/또는 상대적으로 큰 픽셀을 사용함으로써 얻어질 수 있는데, 이는 시야 내 검출된 물체의 짧은 깊이 때문에 프로젝트 이미지에 대한 픽셀의 크기가 상대적으로 크기 때문이다. 이는 비용 절감과 에너지 소비 감소를 가져올 수 있다. 반면, 긴 범위 이미징에 최적화된 센서에 대해, 짧은 범위 이미징에 최적화된 센서에 비해 더 높은 공간 해상도가 사용되어야 한다. 더 높은 공간 해상도는 프레임 내의 상대적으로 많은 픽셀 및/또는 상대적으로 작은 픽셀을 이용함으로써 달성될 수 있는데, 이는 시야 내 검출된 물체의 긴 깊이 때문에 프로젝트 이미지에 대한 픽셀의 크기가 상대적으로 작기 때문이다. 더 높은 공간 해상도는 깊이 측정에 있어 더 높은 정확성을 가져온다.
- [0069] 짧거나 긴 범위 이미징을 최적화하기 위해 설정될 수 있는 센서의 다른 특성은 감도(sensitivity)이다. 감도는 입사 광에 센서가 반응하는 범위를 지칭한다. 감도의 한 가지 척도는 양자 효율성(quantum efficiency)인데, 이는 픽셀과 같은 센서의 광반응(photoreactive) 표면에 입사하여 전자-정공 쌍을 산출하는 광자의 퍼센티지이다. 짧은 범위 이미징에 최적화된 센서에 대해, 더 낮은 감도가 수용가능한데, 이는 센서로 광자를 반사하는 물체의 가까운 거리 때문에 더 많은 광자가 각각의 픽셀에 입사할 것이기 때문이다. 더 낮은 감도는, 예를 들어, 낮은 품질의 센서를 이용하여 달성할 수 있고, 이는 비용 절감 가져온다. 반면, 긴 범위 이미징을 위해 최적화된 센서에 대해, 짧은 거리 이미징에 최적화된 센서에 비해 더 높은 감도가 사용되어야 한다. 센서로 광자를 반사하는 물체의 먼 거리 때문에 상대적으로 더 적은 광자가 각각의 픽셀에 입사하는 경우에 검출을 가능하게 하기 위해, 높은 감도는 고품질 센서를 이용하여 달성될 수 있다.
- [0070] 짧거나 긴 범위 이미징을 최적화하기 위해 설정될 수 있는 센서의 다른 특성은 노출 시간이다. 노출 시간은 이미지 데이터의 프레임을 획득하는 프로세스 동안 센서의 픽셀 상에 광이 떨어지도록 허용되는 시간, 예를 들어, 카메라 셔터가 열리는 시간이다. 노출 시간 동안, 센서의 픽셀은 전하를 축적 또는 집적한다. 더 긴 노출시간

은 낮은 감도를 보상할 수 있다는 점에서 노출 시간은 감도와 관련된다. 그러나, 짧은 범위에서 모션 시퀀스를 정확하게 캡처하기 위해서는 짧은 노출 시간이 바람직한데, 이는 물체가 가까울 때에는 이미징되는 물체의 주어진 움직임이 더 큰 픽셀 오프셋으로 번역되기 때문이다. 짧은 범위 이미징에 최적화된 센서에 대해 짧은 노출 시간이 사용될 수 있는 반면, 긴 범위 이미징에 최적화된 센서에 대해 긴 노출 시간이 사용될 수 있다. 적당한 노출시간을 사용함으로써, 가까운 물체의 과노출(over exposure)/이미지 포화(saturation)와 먼 물체의 과소노출(under exposure)을 피할 수 있다.

[0071] 도 6(b)는 조명기의 일 측에 2개의 센서를 갖고 조명기의 반대측에 하나의 센서를 갖는 이미징 컴포넌트(610)를 도시한다. 이러한 방식으로 제3의 센서를 추가하는 것은 더 적은 오클루전으로 객체를 이미징하게 할 수 있을 뿐만 아니라, 획득되는 추가의 깊이 측정치 때문에 더 정확한 이미징을 가져올 수 있다. 센서(612)와 같은 하나의 센서는 조명기에 가까이 위치될 수 있는 반면, 다른 2개의 센서는 조명기의 반대측에 있다. 이 예에서, 센서(24)는 조명기(26)로부터 베이스라인 거리(BL1)에 있고, 센서(29)는 조명기(26)로부터 베이스라인 거리(BL2)에 있으며, 제3 센서(612)는 조명기(26)로부터 베이스라인 거리(BL3)에 있다.

[0072] 도 6(c)는 조명기의 공통 측에 3개의 센서를 갖는 이미징 컴포넌트(620)를 도시한다. 이러한 방식으로 제3의 센서를 추가하는 것은 획득되는 추가의 깊이 측정치 때문에 더 정확한 이미징을 가져올 수 있다. 또한, 각각의 센서는 상이한 깊이 범위에 대해 최적화될 수 있다. 예를 들어, 조명기로부터 긴 베이스라인 거리(BL3)에서 센서(24)는 긴 범위 이미징에 대해 최적화될 수 있다. 조명기로부터 중간 베이스라인 거리(BL2)에서 센서(29)는 중간 범위 이미징에 대해 최적화될 수 있다. 그리고, 조명기로부터 작은 베이스라인 거리(BL1)에서 센서(612)는 짧은 범위 이미징에 대해 최적화될 수 있다. 유사하게, 공간 해상도, 감도 및/또는 노출 시간은 센서(24)에 대해 긴 범위 레벨, 센서(29)에 대해 중간 범위 레벨, 그리고 센서(612)에 대해 짧은 범위 레벨에 최적화될 수 있다.

[0073] 도 6(d)는 조명기의 반대측에 2개의 센서를 갖는 이미징 컴포넌트(630)를 도시하며, 2개의 센서가 물체의 상이한 부분을 감지하는 방법을 보여준다. 센서 S1(24)는 조명기(26)로부터 베이스라인 거리(BL1)에 있고, 짧은 범위 이미징에 대해 최적화된다. 센서 S2(29)는 조명기(26)로부터 베이스라인 거리(BL2 > BL1)에 있고, 긴 범위 이미징에 대해 최적화된다. RGB 카메라(28)도 도시된다. 물체(660)는 시야에 존재한다. 도면의 관점은 단순화로서, 이미징 컴포넌트(630)는 전면(front view)으로 도시되고 물체(660)는 평면(top view)으로 도시되는 것으로 수정됨을 유의하라. 광선(640 및 642)은 조명기(26)에 의해 프로젝션되는 예시적인 광선이다. 광선(632, 634 및 636)은 센서 S1(24)에 의해 감지되는 예시적 반사광선이고, 광선(650과 652)은 센서 S2(29)에 의해 감지되는 예시적 반사광선이다.

[0074] 물체는 센서 S1(24)와 S2(29)에 의해 감지되는 5개의 표면을 포함한다. 그러나, 오클루전에 의해, 모든 표면이 양 센서에 의해 감지되는 것은 아니다. 예를 들어, 표면(661)은 센서 S1(24)만에 의해 감지되고, 센서 S2(29)의 관점으로부터 가려진다(occluded). 표면(662)도 센서 S1(24)만에 의해 감지되고 센서 S2(29)의 관점으로부터 가려진다. 표면(663)은 센서 S1 및 S2 모두에 의해 감지된다. 표면(664)은 센서 S2만에 의해 감지되고 센서 S1의 관점으로부터 가려진다. 표면(665)은 센서 S2만에 의해 감지되고 센서 S1의 관점으로부터 가려진다. 표면(666)은 센서 S1과 S2 모두에 의해 감지된다. 이는 제2 센서 또는 기타 추가 센서의 부가가 그렇지 않으면 가려졌을 물체의 부분을 이미징하는데 사용될 수 있는 방법을 나타낸다. 또한, 오클루전을 최소화하기 위해 조명기로부터 가능한 멀리 센서를 배치하는 것이 종종 바람직하다.

[0075] 도 7(a)는 시야의 깊이 맵을 획득하기 위한 프로세스를 도시한다. 단계(700)는 구조광의 패턴으로 시야를 조명하는 것을 포함한다. 코딩된 구조광을 포함하는 여하한 타입의 구조광이 사용될 수 있다. 단계 702와 704는 적어도 부분적으로 동시에 수행될 수 있다. 단계 702는 픽셀 데이터의 제1 프레임을 획득하기 위해 제1 센서에서 반사된 적외선 광을 검출하는 것을 포함한다. 이 픽셀 데이터는, 예를 들어, 시야로부터 픽셀에 입사된 광의 양의 표시로서, 노출 시간 동안 각각의 픽셀에 의해 누적된 전하의 양을 나타낼 수 있다. 유사하게, 단계 704는 픽셀 데이터의 제2 프레임을 획득하기 위해 제2 센서에서 반사된 적외선 광을 검출하는 것을 포함한다. 단계 706은 병합된 깊이 맵을 도출하기 위해 모든 프레임으로부터의 픽셀 데이터를 처리하는 것을 포함한다. 이는 도 7(b)-7e와 관련하여 더 설명하는 등의 다른 기술이 관여될 수 있다. 단계 708에서 병합된 깊이 맵에 기초하여 제어 입력을 애플리케이션으로 제공하는 것을 포함한다. 이 제어 입력은 디스플레이 상에서 아바타의 위치를 업데이트하는 것, 사용자 인터페이스(UI)에서 메뉴 아이템을 선택하는 것, 또는 많은 다른 가능한 동작 등의 다양한 목적으로 사용될 수 있다.

[0076] 도 7(b)는 도 7(a)의 단계 706을 더 상세히 도시하는데, 여기서 2개의 구조광 깊이 맵이 병합된다. 이 접근에

서, 제1 및 제2 구조광 깊이 맵이 각각 제1 및 제2 프레임으로부터 획득되고, 2개의 깊이 맵이 병합된다. 프로세스는 2 이상의 깊이 맵 중 여하한 수를 병합하는 것으로 확장될 수 있다. 구체적으로, 단계 720에서, (도 7(a)의 단계 702에서 획득된) 픽셀 데이터의 제1 프레임 내 각각의 픽셀에 대해, 구조광의 패턴을 매칭함으로써 조명 프레임 내 대응 점을 결정하려는 시도가 이루어진다. 일부 경우에, 오클루전 또는 기타 요인에 의해, 조명 프레임 내의 대응 점은 제1 프레임의 하나 이상의 픽셀에 대해 성공적으로 결정되지 않을 수 있다. 단계 722에서, 제1 구조광 깊이 맵이 제공된다. 이 깊이 맵은 제1 프레임 내 각각의 픽셀과 대응하는 깊이 값을 식별할 수 있다. 유사하게, 단계 724에서, (도 7(a)의 단계 704에서 획득된) 픽셀 데이터의 제2 프레임 내 각각의 픽셀에 대해, 조명 프레임 내 대응 점을 결정하려는 시도가 이루어진다. 일부 경우에, 오클루전 또는 기타 요인에 의해, 조명 프레임 내의 대응 점은 제2 프레임의 하나 이상의 픽셀에 대해 성공적으로 결정되지 않을 수 있다. 단계 726에서, 제2 구조광 깊이 맵이 제공된다. 이 깊이 맵은 제2 프레임 내 각각의 픽셀과 대응하는 깊이 값을 식별할 수 있다. 단계 720 및 722는 단계 724 및 726과 적어도 부분적으로 동시에 수행될 수 있다. 단계 728에서, 구조광 깊이 맵이 병합되어 도 7(a)의 단계 706의 병합된 깊이 맵을 도출한다.

[0077] 병합은 비가중 평균, 가중 평균, 정확한 측정 및/또는 신뢰 척도가 관여되는 접근을 포함하는 다른 접근에 기초할 수 있다. 한 가지 접근에서, 각 픽셀에 대해, 2 이상의 깊이 맵 사이에서 깊이 값이 평균된다. 제1 프레임 내 i 번째 픽셀에 대한 깊이 값 $d1$ 및 제2 프레임 내 i 번째 픽셀에 대한 깊이 값 $d2$ 의 예시적인 비가중 평균은 $(d1+d2)/2$ 이다. 제1 프레임 내 i 번째 픽셀에 대한 가중치 $w1$ 의 깊이 값 $d1$ 및 제2 프레임 내 i 번째 픽셀에 대한 가중치 $w2$ 의 깊이 값 $d2$ 의 예시적인 가중 평균은 $(w1 * d1 + w2 * d2) / (w1 + w2)$ 이다. 깊이 값을 병합하는 한 가지 접근은, 센서와 조명기 사이의 베이스라인 거리에 기초하여 프레임의 깊이 값에 가중치를 할당하여, 더 높은 신뢰를 나타내는 더 높은 가중치가 베이스라인 거리가 클 때 할당되고, 더 낮은 신뢰를 나타내는 더 낮은 가중치가 베이스라인 거리가 작을 때 할당된다. 이는 더 큰 베이스라인 거리가 더 정확한 깊이 값을 낳기 때문이다. 예를 들어, 도 6(d)에서, 센서 S1으로부터의 깊이 값에 $w1=BL1/(BL1+BL2)$ 의 가중치를 할당하고 센서 S2로부터의 깊이 값에 $w2=BL2/(BL1+BL2)$ 의 가중치를 할당할 수 있다. 설명하면, $BL=1$ 이고 $BL=2$ 거리 단위라고 가정하면, $w1=1/3$ 이고 $w2=2/3$. 가중치는 픽셀마다 또는 깊이 값마다 적용될 수 있다.

[0078] 상술한 예는 이미지의 센서 S1으로부터의 이미지를 도 6(d)의 거리 $BL1+BL2$ 에 기초하여 센서 S2로부터의 이미지로 스테레오스코픽 매칭하는 것으로부터 획득되는 깊이 값으로 증가될 수 있다. 이 경우, 센서 S1으로부터의 깊이 값에 $w1=BL1/(BL1+BL2+BL1+BL2)$ 를 할당하고, S2로부터의 깊이 값에 가중치 $w2=BL2/(BL1+BL2+BL1+BL2)$ 를 할당하고, S1을 S2로 스테레오스코픽 매칭하는 것으로부터 얻어진 깊이 값에 가중치 $w3=(BL1+BL2)/(BL1+BL2+BL1+BL2)$ 를 할당할 수 있다. 설명하자면, $BL=1$ 이고 $BL=2$ 거리 단위라고 가정하면, $w1=1/6$, $w2=2/6$ 이고 $w3=3/6$ 이다. 추가의 증가에서, 이미지의 센서 S2로부터의 이미지를 도 6(d)의 센서 S1로부터의 이미지로 스테레오스코픽 매칭하는 것으로부터 깊이 값이 획득된다. 이 경우, S1으로부터의 깊이 값에 $w1=BL1/(BL1+BL2+BL1+BL2+BL1+BL2)$ 를 할당하고, S2로부터의 깊이 값에 가중치 $w2=BL2/(BL1+BL2+BL1+BL2+BL1+BL2)$ 를 할당하고, S1을 S2에 스테레오스코픽 매칭하는 것으로부터 획득된 깊이 값에 가중치 $w3=(BL1+BL2)/(BL1+BL2+BL1+BL2+BL1+BL2)$ 를 할당하고, S2를 S1에 스테레오스코픽 매칭하는 것으로부터 얻어진 깊이 값에 가중치 $w4=(BL1+BL2)/(BL1+BL2+BL1+BL2+BL1+BL2)$ 를 할당할 수 있다. 설명하자면, $BL=1$ 이고 $BL=2$ 거리 단위라고 가정하면, $w1=1/9$, $w2=2/9$, $w3=3/9$ 이고 $w4=3/9$ 이다. 이는 하나의 가능성일 뿐이다.

[0079] 가중치는, 높은 신뢰 척도를 갖는 깊이 값이 높은 가중치를 할당받도록 신뢰 척도에 기초하여 제공될 수도 있다. 한 가지 접근에서, 초기 신뢰 척도가 각각의 픽셀에 할당되고, 프레임마다 물체의 깊이가 빨리 변화하지 않는다는 가정에 기초하여, 공차(tolerance) 내에서 깊이 값이 동일하거나 거의 동일한 각각의 새로운 프레임에 대해 신뢰 척도가 증가된다. 예를 들어, 초당 30 프레임의 프레임 레이트에서, 추적되는 사람은 프레임 사이에서 현저하게 움직이지 않을 것이다. 더 상세한 사항에 대해서는, 여기에 참조로 포함되는, 91년 8월 13일 발행 미국 특허 5,040,116, 명칭 "Visual navigation and obstacle avoidance structured light system"을 참조하라. 다른 접근에서, 신뢰 척도는 깊이 값에서의 노이즈 척도이다. 예를 들어, 실제로는 이웃하는 픽셀 사이에서 깊이 값의 큰 변화는 일어나기 어렵다는 가정 하에서, 이러한 깊이 값의 큰 변화는 큰 노이즈를 나타낼 수 있고, 낮은 신뢰 척도를 가져온다. 더 상세한 사항에 대해서는, 여기에 참조로 포함되는, 04년 6월 15일 발행 미국 특허 6,751,338, 명칭 "System and method of using range image data with machine vision tools"을 참조하라. 신뢰 척도를 할당하기 위한 다른 접근도 가능하다.

[0080] 한 가지 접근에서 "마스터" 카메라 좌표계가 정의되고, 다른 깊이 이미지를 "마스터" 좌표계로 변환 및 재샘플링한다(resample). 이미지를 매칭하면, 신뢰를 가중할 수 있는 경우에 하나 이상의 샘플을 고려에 넣도록 선택할 수 있다. 평균은 하나의 해결책이지만, 각 카메라가 공간 내 다른 위치를 성공적으로 관찰하는 경우에 오클

루전을 해결하지 못하므로 반드시 최고의 해결책은 아니다. 신뢰 척도는 깊이 맵 내 각각의 깊이 값에 연관될 수 있다. 다른 접근은 이미지 픽셀이 존재하지 않는 경우에 3D 공간에서 데이터를 병합하는 것이다. 3-D에서, 용적측정(volumetric) 방법이 사용될 수 있다.

[0081] 픽셀이 정확하게 매칭되는 패턴을 갖고 그래서 정확한 깊이 데이터를 갖는지 여부를 결정하기 위해, 보통, 이미지와 알려진 프로젝트된 패턴 사이에서 상관(correlation) 또는 정규화된(normalized) 상관을 수행한다. 이는 센서와 조명기 사이의 에피-폴라 라인을 따라 이루어진다. 성공적인 매칭은 상관의 상대적으로 강한 국대(local maximum)에 의해 표시되는데, 이는 높은 신뢰 척도와 연관될 수 있다. 반면, 상대적으로 약한 상관의 국대는 낮은 신뢰 척도와 연관될 수 있다.

[0082] 가중치는, 높은 정확성 척도를 갖는 깊이 값이 높은 가중치를 할당받도록 정확도 척도에 기초하여서도 제공될 수 있다. 예를 들어, 공간 해상도와 센서와 조명기 사이 및 센서 사이의 베이스라인 거리에 기초하여, 각각의 깊이 샘플에 대해 정확도 척도를 할당할 수 있다. 정확도 척도에 대해 다양한 기술이 알려져 있다. 예를 들어, 캐나다 BC 리치몬드의 Point Grey Research의 "Stereo Accuracy and Error Modeling" (2004.4.19), <http://www.ptgrey.com/support/kb/data/kbStereoAccuracyShort.pdf>를 참조하라. 그 후 이들 정확도에 기초하여 가중 평균을 계산할 수 있다. 예를 들어, 측정된 3D 점에 대해, 가중치 $W_i = \exp(-\text{accuracy}_i)$ 를 할당하는데, 여기서, accuracy_i 는 정확도 척도이고, 평균된 3D 점은 $P_{\text{avg}} = \sum(W_i * P_i) / \sum(W_i)$ 이다. 그러면, 이들 가중치를 이용하여, 3-D에서 근접한 점 샘플이 가중 평균을 이용하여 병합될 수 있다.

[0083] 3D에서 깊이 값 데이터를 병합하기 위해, $(X, Y, Z) = \text{깊이} * \text{광선} + \text{원점}$ 을 이용하여 모든 깊이 이미지를 3D 공간에 프로젝션할 수 있는데, 여기서 광선은 픽셀로부터 센서의 초점까지의 3D 벡터이고, 원점은 3D 공간에서 센서의 초점의 위치이다. 3D 공간에서, 각각의 깊이 데이터 점에 대해 법선(normal) 방향을 계산한다. 또한, 각각의 데이터 점에 대해, 다른 소스로부터 근처 데이터 점을 찾는다. 다른 데이터 점이 충분히 가깝고 점의 법선 벡터들 사이의 내적(dot product)이 양(positive)이면, 이는 이들이 유사한 방향을 갖고 물체의 2개 측면이 아님을 의미하고, 이 점들을 단일의 점으로 병합한다. 이 병합은, 예를 들어, 점의 3D 위치의 가중 평균을 계산함으로써 수행될 수 있다. 가중치는 측정치의 신뢰에 의해 정의될 수 있고, 여기서 신뢰 척도는 상관 스코어에 기초할 수 있다.

[0084] 도 7(c)는 도 7(a)의 단계 706을 더 상세히 도시하고, 여기서 2개의 구조광 깊이 맵 및 2개의 스테레오스코픽 깊이 맵이 병합된다. 이 접근에서, 제1 및 제2 구조광 깊이 맵이 제1 및 제2 프레임 각각으로부터 획득된다. 추가적으로, 하나 이상의 스테레오스코픽 깊이 맵이 획득된다. 제1 및 제2 구조광 깊이 맵과 하나 이상의 스테레오스코픽 깊이 맵이 병합된다. 프로세스는 2 이상의 깊이 맵 중 여하한 갯수를 병합하는데까지 확장될 수 있다. 단계 740과 742는 단계 744와 746, 단계 748과 750, 및 단계 752와 754와 적어도 부분적으로 동시에 수행될 수 있다. 단계 740에서, 픽셀 데이터의 제1 프레임 내 각각의 픽셀에 대해, 조명 프레임 내 대응 점을 결정하고, 단계 742에서 제1 구조광 깊이 맵을 제공한다. 단계 744에서, 픽셀 데이터의 제1 프레임 내 각각의 픽셀에 대해, 픽셀 데이터의 제2 프레임 내의 대응 픽셀을 결정하고, 단계 746에서 제1 스테레오스코픽 깊이 맵을 제공한다. 단계 748에서, 픽셀 데이터의 제2 프레임 내 각각의 픽셀에 대해, 조명 프레임 내 대응 점을 결정하고, 단계 750에서 제2 구조광 깊이 맵을 제공한다. 단계 752에서, 픽셀 데이터의 제2 프레임 내 각각의 픽셀에 대해, 픽셀 데이터의 제1 프레임 내의 대응 픽셀을 결정하고, 단계 754에서 제2 스테레오스코픽 깊이 맵을 제공한다. 단계 756은 상이한 깊이 맵들을 병합하는 것을 포함한다.

[0085] 병합은, 비가중 평균, 가중 평균, 정확도 척도 및/또는 신뢰 척도가 관여되는 접근을 포함하는 상이한 접근에 기초할 수 있다.

[0086] 이 접근에서, 2개의 스테레오스코픽 깊이 맵이 2개의 구조광 깊이 맵과 병합된다. 한 가지 옵션에서, 병합은 단일의 병합 단계에서 모든 깊이 맵을 함께 고려한다. 다른 가능한 접근에서, 병합은 다수의 단계에서 일어난다. 예를 들어, 구조광 깊이 맵은 제1 병합 깊이 맵을 획득하기 위해 병합될 수 있고, 스테레오스코픽 깊이 맵은 제2 병합 깊이 맵을 획득하기 위해 병합될 수 있으며, 제1 및 제2 병합 깊이 맵은 최종 병합 깊이 맵을 획득하기 위해 병합된다. 병합이 다수의 단계에서 일어나는 다른 옵션에서, 제1 병합 깊이 맵을 획득하기 위해 제1 구조광 깊이 맵이 제1 스테레오스코픽 깊이 맵과 병합되고, 제2 병합 깊이 맵을 획득하기 위해 제2 구조광 깊이 맵이 제2 스테레오스코픽 깊이 맵과 병합되고, 최종 병합 깊이 맵을 획득하기 위해 제1 및 제2 병합 깊이 맵이 병합된다. 다른 접근도 가능하다.

[0087] 다른 접근에서, 단 하나의 스테레오스코픽 깊이 맵이 2개의 구조광 깊이 맵과 병합된다. 하나 이상의 단계에서 병합이 일어날 수 있다. 다단계 접근에서, 제1 병합 깊이 맵을 획득하기 위해 제1 구조광 깊이 맵이 스테레오

스코픽 깊이 맵과 병합되고, 최종 병합 깊이 맵을 획득하기 위해 제2 구조광 깊이 맵이 스테레오스코픽 깊이 맵과 병합된다. 또는, 제1 병합 깊이 맵을 획득하기 위해 2개의 구조광 깊이 맵이 병합되고, 최종 병합 깊이 맵을 획득하기 위해 제1 병합 깊이 맵이 스테레오스코픽 깊이 맵과 병합된다. 다른 접근이 가능하다.

[0088] 도 7(d)는 도 7(a)의 단계 706을 더 상세히 도시하는데, 여기서 깊이 값이 필요에 따라 스테레오스코픽 매칭을 이용하여 개선(refine)된다. 이 접근은 개선이 바람직함을 나타내는 조건을 검출하는데 응답하여 하나 이상의 깊이 값을 개선하기 위해 스테레오스코픽 매칭이 사용된다는 점에서 적응적(adaptive)이다. 스테레오스코픽 매칭은 프레임 내 픽셀의 부분집합에 대해서만 수행될 수 있다. 한 가지 접근에서, 픽셀이 구조광 패턴과 매칭될 수 없어 깊이 값이 널(null) 또는 디폴트 값일 때에 픽셀의 깊이 값의 개선이 바람직하다. 픽셀은 오클루전, 음영(shadowing), 광 조건, 표면 질감 또는 기타 이유로 구조광 패턴에 매칭되지 않을 수 있다. 이 경우, 깊이 값이 이전에 획득되지 않은 경우에는 스테레오스코픽 매칭이 깊이 값을 제공할 수 있고, 또는 일부 경우에, 센서와 조명기 사이의 베이스라인 간격에 비해 더 큰 베이스라인으로 센서가 떨어져 있기 때문에 더 정확한 깊이 값을 제공할 수 있다. 예를 들어, 도 2, 6b 및 6d를 참조하라.

[0089] 다른 접근에서, 깊이 값이 문턱 거리를 초과하여 물체의 대응 점이 상대적으로 센서로부터 멀다는 것을 나타내는 때에, 픽셀의 깊이 값의 개선이 바람직하다. 이 경우, 센서 사이의 베이스라인이 각각의 센서와 조명기 사이의 베이스라인보다 큰 경우에 스테레오스코픽 매칭이 더 정확한 깊이 값을 제공할 수 있다.

[0090] 개선에는, 이전에 깊이 값이 제공되지 않은 경우 깊이 값을 제공하는 것, 또는, 예를 들어, 비가중 평균, 가중 평균, 정확도 척도 및/또는 신뢰 척도가 관여되는 다른 접근에 기초하여 깊이 값을 병합하는 것이 관여될 수 있다. 또한, 개선은 깊이 값이 병합되기 전에 각각의 센서의 프레임에 대해 별도로 수행될 수 있다.

[0091] 개선이 바람직함을 나타내는 조건이 검출되는 픽셀에 대해서만 스테레오스코픽 매칭을 수행함으로써, 불필요한 처리를 피한다. 스테레오스코픽 매칭은 개선이 바람직함을 나타내는 조건이 검출되지 않은 픽셀에 대해 수행되지 않는다. 그러나, 프레임 내 하나 이상의 픽셀에 대해 개선이 바람직함을 나타내는 조건이 검출되는 때에 전체 프레임에 대해 스테레오스코픽 매칭을 수행하는 것도 가능하다. 한 가지 접근에서, 프레임 내 픽셀의 일부 중 최소 개수에 대해 개선이 표시되는 때에 전체 프레임에 대한 스테레오스코픽 매칭이 시작된다.

[0092] 단계 760에서, 픽셀 데이터의 제1 프레임 내 각각의 픽셀에 대해, 조명 프레임 내의 대응 점을 결정하고, 단계 761에서, 대응하는 제1 구조광 깊이 맵을 제공한다. 결정 단계 762는 깊이 값의 개선이 표시되는지 여부를 결정한다. 기준은, 픽셀 데이터의 제1 프레임 내 각각의 픽셀에 대해 평가될 수 있고, 한 가지 접근에서, 픽셀과 연관된 깊이 값의 개선이 바람직한지 여부를 표시할 수 있다. 한 가지 접근에서, 관련 깊이 값이 사용가능하지 않거나 신뢰할 수 없을 때 개선이 바람직하다. 신뢰할 수 없음은 예를 들어 정확도 척도 및/또는 신뢰 척도에 기초할 수 있다. 신뢰 척도가 문턱 신뢰 척도를 초과하면, 깊이 값은 신뢰할 수 있는 것으로 간주될 수 있다. 또는, 정확도 척도가 문턱 정확도 척도를 초과하면, 깊이 값은 신뢰할 수 있는 것으로 간주될 수 있다. 다른 접근에서, 깊이 값이 신뢰할 수 있는 것으로 간주되기 위해 신뢰 척도 및 정확도 척도 모두가 각각의 문턱 수준을 초과하여야 한다.

[0093] 다른 접근에서, 깊이가 문턱 깊이를 넘는 때와 같이, 관련 깊이 값이 깊이가 상대적으로 멀다는 것을 나타내는 때에 개선이 바람직하다. 개선이 바람직하면, 단계 763는 픽셀 데이터의 제2 프레임 내 하나 이상의 픽셀에 대한 픽셀 데이터의 제1 프레임 내 하나 이상의 픽셀의 스테레오스코픽 매칭을 수행한다. 이는 픽셀 데이터의 제1 프레임의 하나 이상의 추가 깊이 값을 가져온다.

[0094] 유사하게, 픽셀 데이터의 제2 프레임에 대해, 단계 764에서, 픽셀 데이터의 제2 프레임 내 각각의 픽셀에 대해, 조명 프레임 내의 대응 점을 결정하고, 단계 765에서, 대응하는 제2 구조광 깊이 맵을 제공한다. 결정 단계 766은 깊이 값의 개선이 표시되는지 여부를 결정한다. 개선이 바람직하면, 단계 767은 픽셀 데이터의 제1 프레임 내 하나 이상의 픽셀에 대한 픽셀 데이터의 제2 프레임 내 하나 이상의 픽셀의 스테레오스코픽 매칭을 수행한다. 이는 픽셀 데이터의 제2 프레임의 하나 이상의 추가 깊이 값을 가져온다.

[0095] 단계 768은 픽셀 데이터의 제1 및 제2 프레임의 깊이 맵을 병합하는데, 여기서 병합은 단계 763 및/또는 767의 스테레오스코픽 매칭으로부터 획득되는 깊이 값을 포함한다. 병합은 비가중 평균, 가중 평균, 정확도 척도 및/또는 신뢰 척도가 관여되는 접근을 포함하여 다른 접근에 기초할 수 있다.

[0096] 개선이 표시된 주어진 픽셀에 대해, 병합은 제1 구조광 깊이 맵으로부터의 깊이 값, 제2 구조광 깊이 맵으로부터의 깊이 값, 그리고 스테레오스코픽 매칭으로부터의 하나 이상의 깊이 값을 병합할 수 있음을 유의하라. 이 접근은 구조광 깊이 맵으로부터의 깊이 값을 폐기하고 스테레오스코픽 매칭으로부터의 깊이 값으로 대체하는 접

근에 비해 더 신뢰할 수 있는 결과를 제공할 수 있다.

- [0097] 도 7(e)는 도 7(a)의 단계 706에 대한 다른 접근을 더 상세히 도시하는데, 여기서 병합된 깊이 맵의 깊이 값은 필요에 따라 스테레오스코픽 매칭을 이용하여 개선된다. 이 접근에서, 구조광 패턴에 매칭함으로써 획득되는 깊이 맵의 병합은 개선 프로세스 전에 일어난다. 단계 760, 761, 764 및 765는 도 7(d)의 동일 번호의 단계와 동일하다. 단계 770는 구조광 깊이 맵을 병합한다. 병합은, 비가중 평균, 가중 평균, 정확도 척도 및/또는 신뢰도 척도가 관여되는 접근을 포함하여 다른 접근에 기초할 수 있다. 단계 771은 도 7(d)의 단계 762 및 766과 같고, 깊이 값의 개선이 표시되는지 여부를 결정하는 것을 포함한다.
- [0098] 기준은, 픽셀 데이터의 병합 깊이 맵 내 각각의 픽셀에 대해 평가될 수 있고, 한 가지 접근에서, 픽셀과 연관된 깊이 값의 개선이 바람직한지 여부를 표시할 수 있다. 한 가지 접근에서, 관련 깊이 값이 사용가능하지 않거나 신뢰할 수 없을 때 개선이 바람직하다. 신뢰할 수 없음은 예를 들어 정확도 척도 및/또는 신뢰 척도에 기초할 수 있다. 신뢰 척도가 문턱 신뢰 척도를 초과하면, 깊이 값은 신뢰할 수 있는 것으로 간주될 수 있다. 또는, 정확도 척도가 문턱 정확도 척도를 초과하면, 깊이 값은 신뢰할 수 있는 것으로 간주될 수 있다. 다른 접근에서, 깊이 값이 신뢰할 수 있는 것으로 간주되기 위해 신뢰 척도 및 정확도 척도 모두가 각각의 문턱 수준을 초과하여야 한다. 다른 접근에서, 깊이가 문턱 깊이를 넘는 때와 같이, 관련 깊이 값이 깊이가 상대적으로 멀다는 것을 나타내는 때에 개선이 바람직하다. 개선이 바람직하면, 단계 772 및/또는 773이 수행될 수 있다. 일부 경우에, 한 프레임 내 픽셀을 다른 프레임 내 픽셀에 매칭함으로써 한 방향의 스테레오스코픽 매칭이 수행되는 것으로 충분하다. 다른 경우에, 양 방향으로 스테레오스코픽 매칭이 수행될 수 있다. 단계 772는 픽셀 데이터의 제2 프레임 내 하나 이상의 픽셀에 대한 픽셀 데이터의 제1 프레임 내 하나 이상의 픽셀의 스테레오스코픽 매칭을 수행한다. 이는 픽셀 데이터의 제1 프레임의 하나 이상의 추가 깊이 값을 가져온다. 단계 773은 픽셀 데이터의 제1 프레임 내 하나 이상의 픽셀에 대한 픽셀 데이터의 제2 프레임 내 하나 이상의 픽셀의 스테레오스코픽 매칭을 수행한다. 이는 픽셀 데이터의 제2 프레임의 하나 이상의 추가 깊이 값을 가져온다.
- [0099] 단계 774는 스테레오스코픽 매칭이 수행된 하나 이상의 선택된 픽셀에 대해 단계 770의 병합된 깊이 맵을 개선한다. 개선에는, 비가중 평균, 가중 평균, 정확도 척도 및/또는 신뢰도 척도가 관여되는 접근을 포함하여 다른 접근에 기초한 깊이 값 병합이 관여될 수 있다.
- [0100] 단계 771에서 개선이 바람직하지 않으면, 프로세스는 단계 775에서 끝난다.
- [0101] 도 8은 도 7(a)의 단계 708에 제시된 제어 입력을 이용하여 사람 타겟을 추적하기 위한 예시적인 방법을 도시한다. 언급된 바와 같이, 깊이 카메라 시스템은 제스처와 같은 사용자의 움직임을 추적하는데 사용될 수 있다. 움직임은 애플리케이션에서 제어 입력으로서 처리될 수 있다. 예를 들어, 이는 도 1에 도시된 바와 같이 아바타가 사용자를 나타내는 경우에 디스플레이 상의 아바타의 위치를 업데이트하는 것, 사용자 인터페이스(UI)에서 메뉴 아이템을 선택하는 것 또는 기타 많은 가능한 동작을 포함할 수 있다.
- [0102] 예시적인 방법은, 예를 들어, 도 2-4와 관련하여 논의된 바와 같은 깊이 카메라 시스템(20) 및/또는 컴퓨팅 환경(12, 100 또는 420)을 이용하여 구현될 수 있다. 골격 모델, 메시 휴먼(mesh human) 모델 또는 기타 여하한 사람의 적당한 표현과 같은 모델을 생성하기 위해 하나 이상의 사람 타겟이 스캐닝될 수 있다. 골격 모델에서, 각각의 신체 부분은 골격 모델의 관절과 골격을 정의하는 수학적 벡터로 특징지어질 수 있다. 신체 부분은 관절에서 서로에 대해 움직일 수 있다.
- [0103] 그러면 모델은 컴퓨팅 환경에 의해 실행되는 애플리케이션과 상호작용하는데 사용될 수 있다. 모델을 생성하기 위한 스캐닝은 애플리케이션이 시작 또는 론칭되는 때에, 또는 스캐닝되는 사람의 애플리케이션에 의해 제어되는 다른 시간에 일어날 수 있다.
- [0104] 사용자의 물리적 움직임 또는 모션이 애플리케이션의 파라미터를 조정 및/또는 제어하는 실시간 사용자 인터페이스로서 동작하도록 추적될 수 있는 골격 모델을 생성하기 위해 사람이 스캐닝될 수 있다. 예를 들어, 사람의 추적된 움직임은, 전자 롤-플레이 게임에서 아바타나 기타 스크린 상의 캐릭터를 움직이거나, 전자 레이싱 게임에서 스크린 상의 차량을 제어하거나, 가상 환경에서 건물이나 물체의 조적을 제어하거나, 애플리케이션의 여하한 기타 적당한 제어를 수행하는데 사용될 수 있다.
- [0105] 일 실시형태에 따르면, 단계 800에서, 예를 들어 깊이 카메라 시스템으로부터 깊이 정보가 수신된다. 깊이 카메라 시스템은 하나 이상의 타겟을 포함할 수 있는 시야를 캡처 또는 관찰할 수 있다. 깊이 정보는 복수의 관찰된 픽셀을 갖는 깊이 이미지 또는 맵을 포함할 수 있는데, 여기서 각각의 관찰된 픽셀은 논의된 바와 같이 관찰된 깊이 값을 갖는다.

- [0106] 깊이가 이미지는 더 낮은 처리 해상도로 다운샘플링될 수 있어서 더 쉽게 사용되고 더 적은 연산 오버헤드로 처리될 수 있다. 또한, 하나 이상의 고분산(high-variance) 및/또는 잡음이 많은(noisy) 깊이 값이 깊이 이미지로부터 제거되거나 및/또는 평판화(smooth)될 수 있고; 누락 및/또는 제거된 깊이 정보의 부분이 채워지거나 및/또는 재구성될 수 있고; 및/또는 깊이 정보가 골격 모델(도 9 참조)과 같은 모델을 생성하는데 사용될 수 있도록 여하한 기타 적당한 처리가 수신된 깊이 정보에 대해 수행될 수 있다.
- [0107] 단계 802는 사람 타겟이 깊이 이미지를 포함하는지 여부를 결정한다. 이는 깊이 이미지가 사람 타겟을 포함하는지를 판정하기 위해 각각의 타겟 또는 물체를 패턴에 비교하여 깊이 이미지 내 각각의 타겟이나 물체를 플러드 필링(flood filling)하는 것을 포함할 수 있다. 예를 들어, 상술한 바와 같이 타겟 또는 물체를 정의할 수 있는 에지(edge)를 결정하기 위해 선택된 영역 내의 픽셀의 다양한 깊이 값 또는 깊이 이미지의 점이 비교될 수 있다. Z 층의 가능한 Z 값이 결정된 에지에 기초하여 플러드 필링될 수 있다. 예를 들어, 결정된 에지와 관련된 픽셀 및 에지 내 영역의 픽셀은, 패턴과 비교될 수 있는 캡처된 영역 내 타겟 또는 물체를 정의하기 위해 서로 연관될 수 있는데, 이는 아래에서 더 상세히 설명한다.
- [0108] 결정 단계 804에서 깊이 이미지가 사람 타겟을 포함하면, 단계 806이 수행된다. 결정 단계 804가 부(false)이면, 추가의 깊이 정보가 단계 800에서 수신된다.
- [0109] 각각의 타겟 또는 물체가 비교되는 패턴은, 총체적으로 보통의 신체를 정의하는 변수의 세트를 갖는 하나 이상의 데이터 구조를 포함할 수 있다. 예를 들어, 시야 내의 사람 타겟 및 사람이 아닌 타겟(non-human target)의 픽셀과 연관된 정보는 사람 타겟을 식별하기 위해 변수와 비교될 수 있다. 일 실시형태에서, 세트 내의 각각의 변수는 신체 부분에 기초하여 가중될 수 있다. 예를 들어, 패턴 내의 머리 및/또는 어깨와 같은 다양한 신체 부분은 다리와 같은 같은 신체 부분에 비해 더 클 수 있는 연관 가중치를 가질 수 있다. 일 실시형태에 따르면, 타겟이 사람인지 여부와 어떤 타겟이 사람인지를 결정하기 위해 타겟을 변수와 비교하는 때에 가중치가 사용될 수 있다. 예를 들어, 큰 가중치를 갖는 타겟 및 변수 사이의 매칭은 작은 가중치를 갖는 매칭에 비해 타겟이 사람일 가능성이 더 높다.
- [0110] 단계 806은 신체 부분에 대해 사람 타겟을 스캐닝하는 것을 포함한다. 사람의 정확한 모델을 제공하기 위해, 사람 타겟은 사람의 하나 이상의 신체 부분과 연관된 길이, 폭 등과 같은 척도를 제공하기 위해 스캐닝될 수 있다. 예시적인 실시형태에서, 사람 타겟이 고립(isolate)될 수 있고 하나 이상의 신체 부분에 대한 스캐닝을 위해 사람 타겟의 비트마스크(bitmask)가 생성될 수 있다. 비트마스크는, 예를 들어, 사람 타겟이 캡처 영역 요소 내 다른 타겟이나 물체와 분리될 수 있도록 사람 타겟을 플러드 필링함으로써 생성될 수 있다. 그러면 비트마스크는 사람 타겟의 골격 모델, 메시 휴먼 모델 등과 같은 모델을 생성하기 위해 하나 이상의 신체 부분에 대하여 분석될 수 있다. 예를 들어, 일 실시형태에 따르면, 스캐닝된 비트마스크에 의해 결정된 측정 값은 골격 모델에서 하나 이상의 관절을 정의하기 위해 사용될 수 있다. 하나 이상의 관절이 사람의 신체 부분에 대응할 수 있는 하나 이상의 뼈를 정의하는데 사용될 수 있다.
- [0111] 예를 들어, 사람 타겟의 비트마스크의 정점(top)은 정수리(top of head) 위치에 연관될 수 있다. 정수리를 결정한 후에, 비트마스크가 아래쪽으로 스캐닝되고 그래서 목의 위치, 어깨의 위치 등을 결정할 수 있다. 예를 들어, 스캐닝되고 있는 위치에서, 비트마스크의 폭은, 예를 들어, 목, 어깨 등과 연관되는 통상의 폭의 문턱 값에 비교될 수 있다. 다른 실시형태에서, 스캐닝되고 비트마스크 내 신체 부분과 연관된 이전 위치로부터의 거리가 목, 어깨 등의 위치를 결정하는데 사용될 수 있다. 다리, 발 등의 일부 신체 부분은, 예를 들어, 다른 신체 부분의 위치에 기초하여 계산될 수 있다. 신체 부분의 값을 결정하는 때에, 신체 부분의 측정 값을 포함하는 데이터 구조가 생성된다. 데이터 구조는 깊이 카메라 시스템에 의해 다른 점에서 제공되는 다수의 깊이 이미지로부터 평균된 스캐닝 결과를 포함할 수 있다.
- [0112] 단계 808은 사람 타겟의 모델을 생성하는 것을 포함한다. 일 실시형태에서, 스캐닝된 비트마스크에 의해 결정되는 측정 값은 골격 모델에서 하나 이상의 관절을 정의하는데 사용될 수 있다. 하나 이상의 관절은 사람의 신체 부분에 대응하는 하나 이상의 뼈를 정의하는데 사용될 수 있다.
- [0113] 더 정확한 골격 모델을 생성하기 위하여 관절이 사람의 신체 부분과 관절 사이의 통상의 거리 범위 내가 될 때까지 하나 이상의 관절이 조정될 수 있다. 모델은, 예를 들어, 사람 타겟과 연관된 키에 기초하여 더 조정될 수 있다.
- [0114] 단계 810에서, 모델은 사람의 위치를 초당 수회 업데이트함으로써 추적된다. 사용자가 물리적 공간에서 움직임에 따라, 골격 모델이 사람을 나타내도록 골격 모델을 조정하기 위해 깊이 카메라 시스템으로부터의 정보가 사

용된다. 특히, 물리적 공간에서의 사람 타겟의 포즈에 더 가깝게 대응하는 포즈로 골격 모델을 조정하기 위해 골격 모델의 하나 이상의 수력(force-receiving) 측면에 하나 이상의 힘이 가해질 수 있다.

[0115] 일반적으로, 사람의 움직임을 추적하기 위한 여하한 알려진 기술이 사용될 수 있다.

[0116] 도 9는 도 8의 단계 808에서 제시된 사람 타겟의 예시적 모델을 도시한다. 모델(900)은 도 1의 z 방향으로 깊이 카메라를 마주하고 있어서, 도시된 단면은 x-y 평면에 있다. 모델은, 예를 들어, 정수리(902), 머리 아래 또는 턱(913), 오른쪽 어깨(904), 오른쪽 팔꿈치(906), 오른쪽 손목(908) 및 손끝 영역으로 표시되는 오른손(910)과 같은 다수의 기준점을 포함한다. 오른쪽과 왼쪽은 카메라를 바라보는 사용자의 입장에서 정의된다. 모델은 또한 왼쪽 어깨(914), 왼쪽 팔꿈치(916), 왼쪽 손목(918) 및 왼손(920)을 포함한다. 허리 영역(922)도, 오른쪽 엉덩이(924), 오른쪽 무릎(926), 오른발(928), 왼쪽 엉덩이(930), 왼쪽 무릎(932) 및 왼발(934)과 함께 도시된다. 어깨선(912)은 어깨(904와 914) 사이의 통상 수평인 선이다. 점(922와 913) 사이에서 연장하는 상체 중심선(925)도 도시된다.

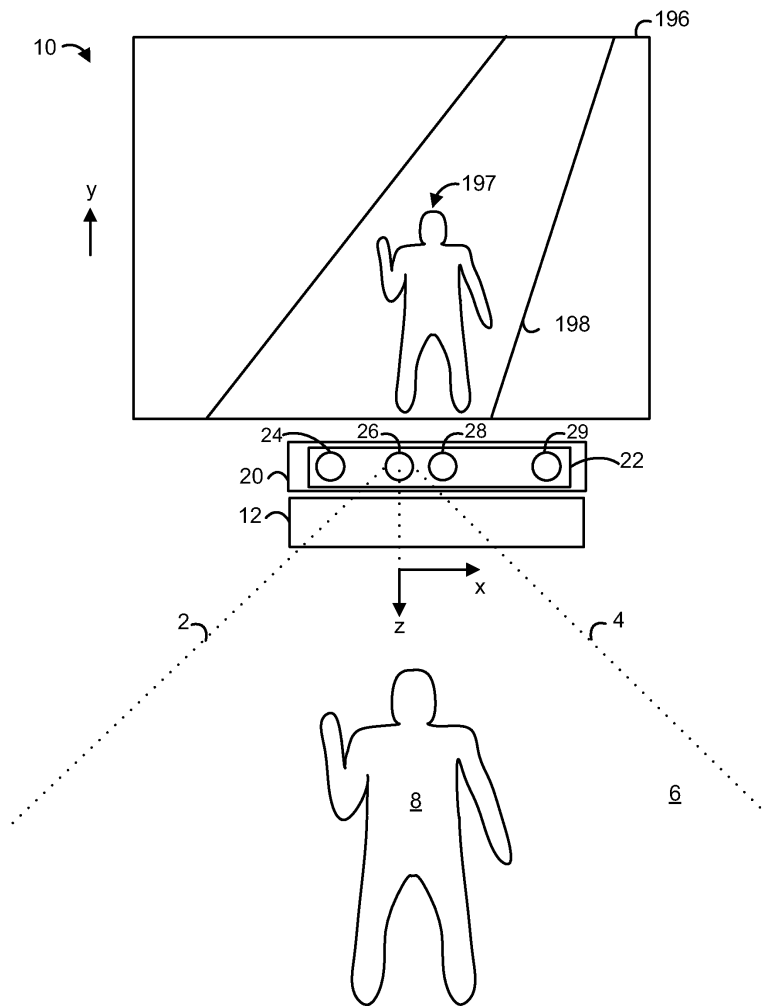
[0117] 따라서, 많은 장점을 갖는 깊이 카메라 시스템이 제공됨을 알 수 있다. 하나의 장점은 오클루전 감소이다. 더 넓은 베이스라인이 사용되므로, 하나의 센서는 다른 센서에게 가려진 정보를 볼 수 있다. 2개의 깊이 맵을 융합하는 것은 단일 센서에 의해 산출되는 맵에 비해 더 관찰가능한 물체를 갖는 3D 이미지를 산출한다. 다른 장점은 감소된 음영 효과이다. 구조광 방법은 내재적으로 센서에게는 보이지만 광원에게는 보이지 않는 위치에서 음영 효과를 만든다. 이들 영역에서 스테레오스코픽 매칭을 적용함으로써, 이 효과가 감소될 수 있다. 다른 장점은 외부 광에 대한 강인성이다. 외부 광이 구조광 카메라를 혼란시켜서 유효한 결과를 산출할 수 없는 많은 시나리오가 있다. 이들 경우에, 스테레오스코픽 데이터가 추가의 척도로서 획득되는데, 실제로 외부 광이 거리를 측정하는데 있어서 이를 도울 수 있기 때문이다. 외부 광은 동일한 장면에서 바라보는 동일한 카메라로부터 올 수 있음을 유의하라. 달리 말하면, 동일한 장면을 바라보는 2 이상의 제안된 카메라를 동작하는 것이 가능하다. 이는, 하나의 카메라에 의해 산출된 광 패턴이 다른 카메라가 패턴을 적절하게 매칭하는데 혼란을 줄 수 있더라도, 여전히 스테레오스코픽 매칭은 성공할 가능성이 높기 때문이다. 다른 장점은, 제안된 구성을 이용하여, 2개의 센서가 더 넓은 베이스라인을 갖는다는 사실 때문에 먼 거리에서 더 높은 정확도를 달성할 수 있다는 것이다. 구조광과 스테레오스코픽 측정 정확도는 모두 센서/프로젝터 사이의 거리에 크게 의존한다.

[0118] 본 기술의 이상의 상세한 설명은 예시와 설명의 목적으로 제시되었다. 이는 망라적인 것이 아니고 정확히 개시된 형태로 본 기술을 제한하려는 것도 아니다. 위의 교시에 비추어 많은 수정과 변경이 가능하다. 설명된 실시형태는, 본 기술의 원리 및 그의 실제적 적용을 가장 잘 설명하여 당업자가 생각하는 구체적인 용도에 맞도록 다양한 실시형태에서 다양한 변경으로 본 기술을 가장 잘 이용할 수 있게 하기 위해서 선택되었다. 본 기술의 범위는 첨부된 청구범위에 의해 정의되도록 한 것이다.

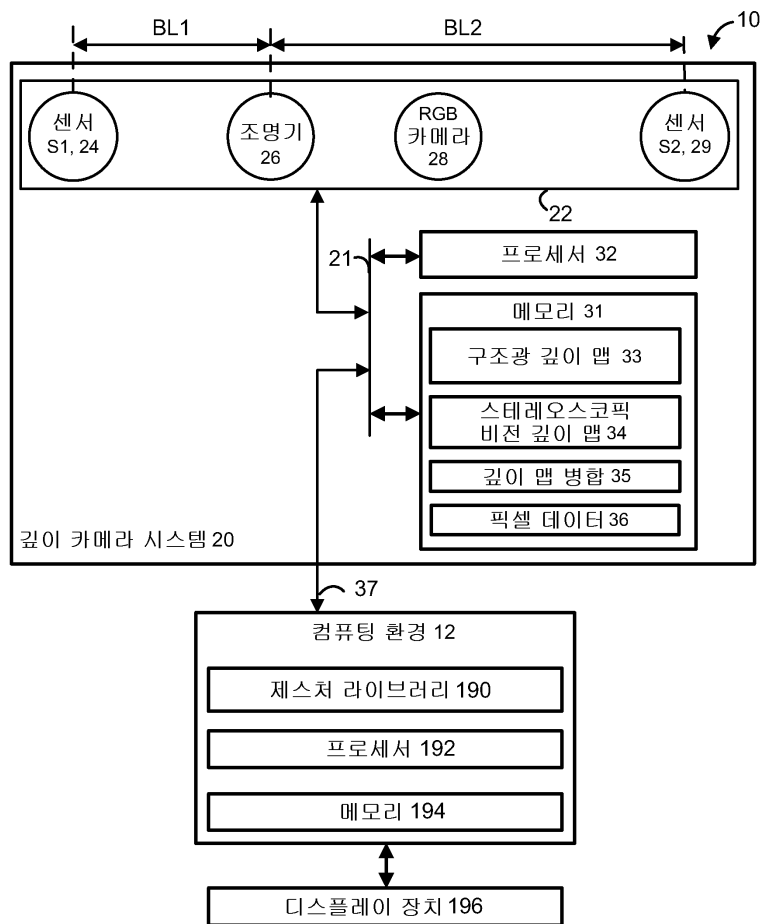
[0119]

도면

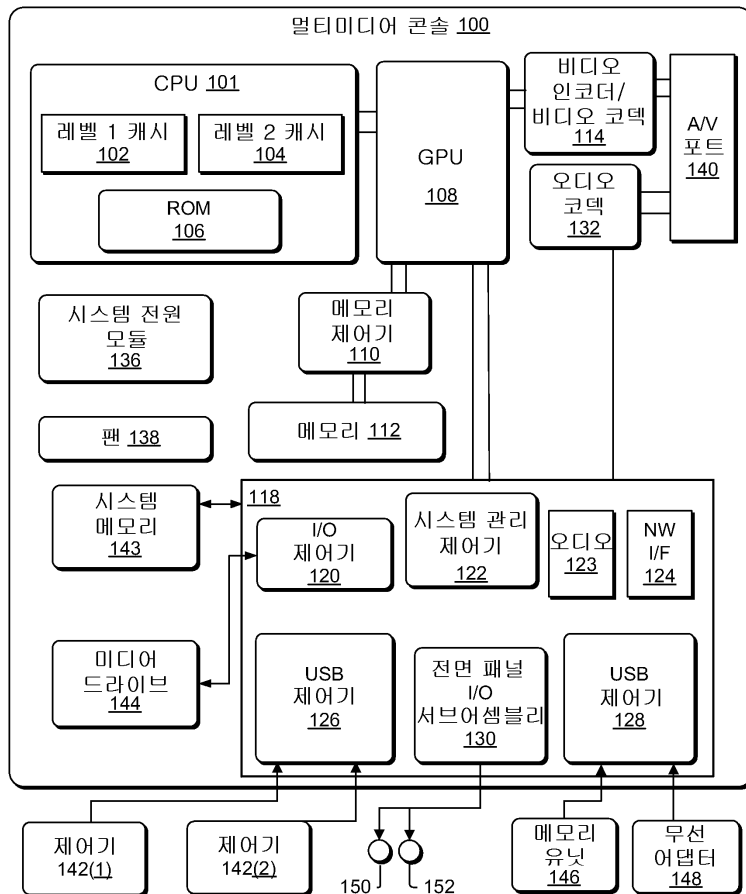
도면1



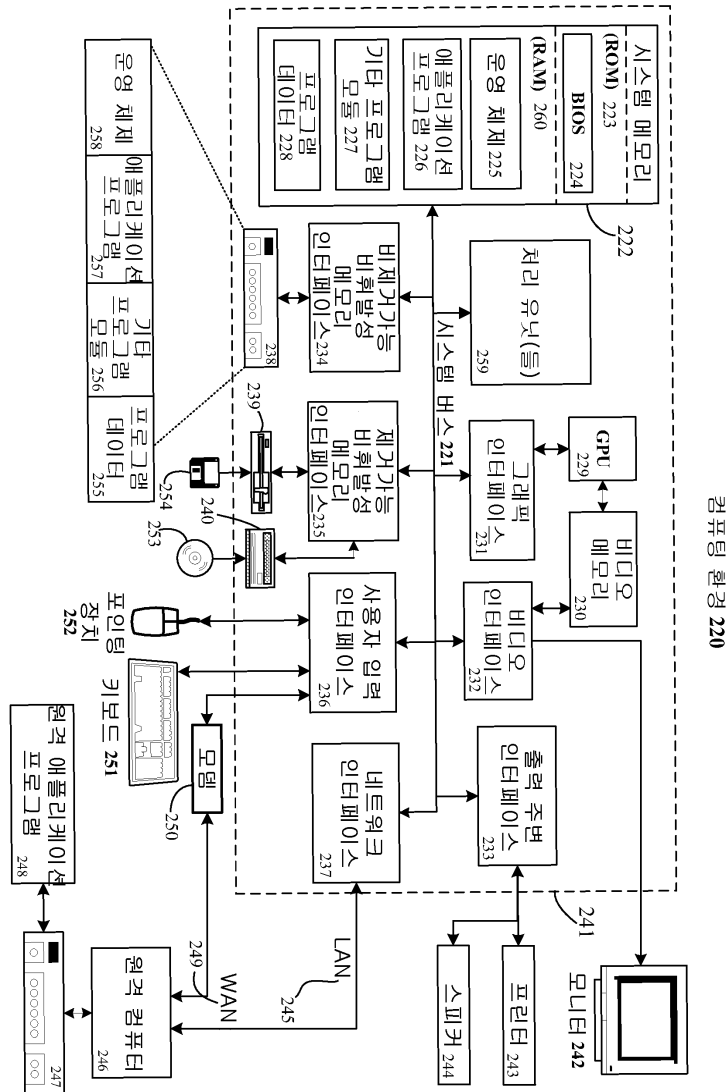
도면2



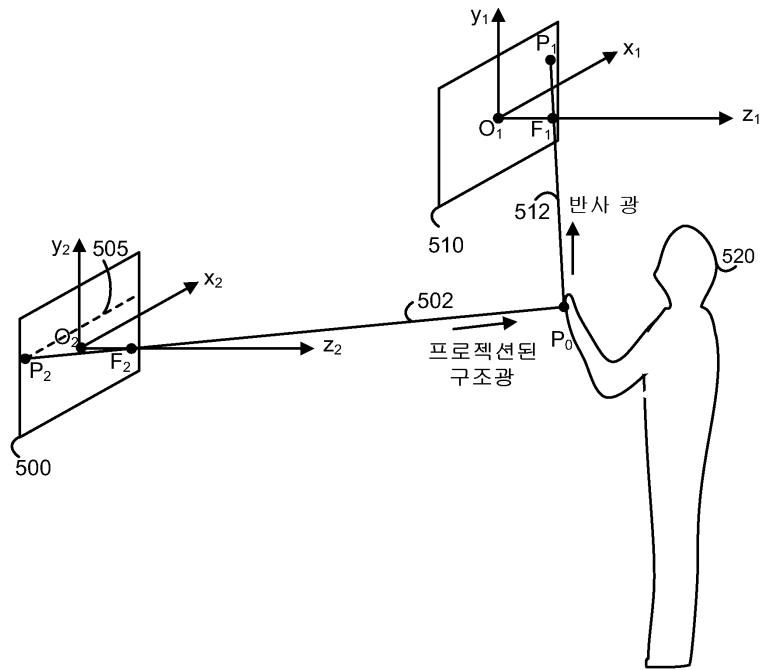
도면3



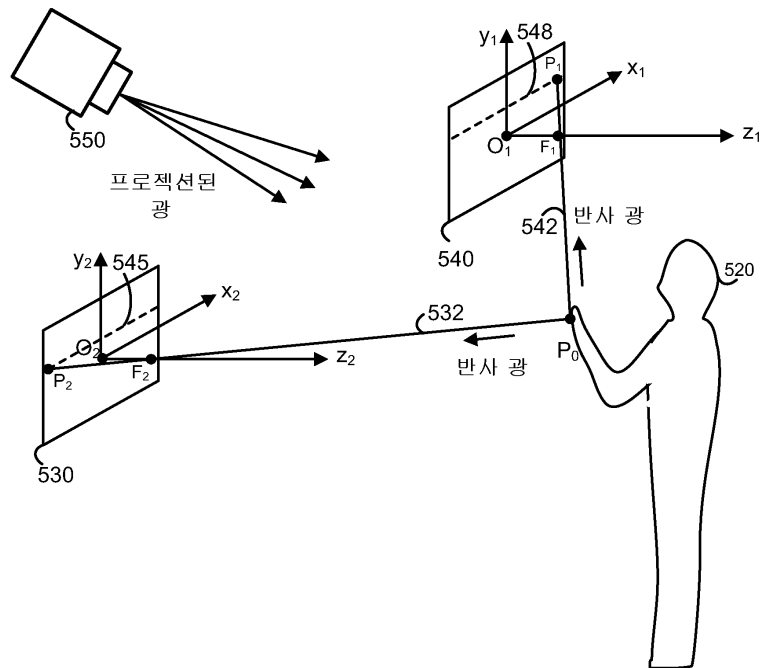
도면4



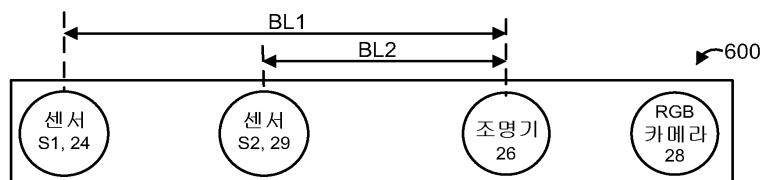
도면5a



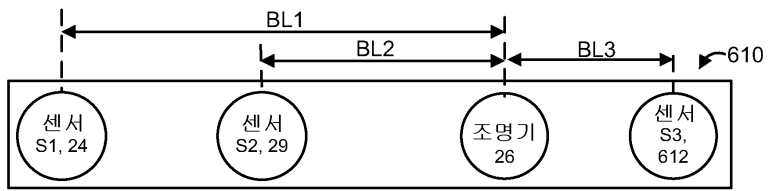
도면5b



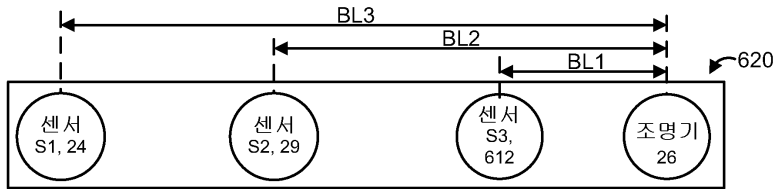
도면6a



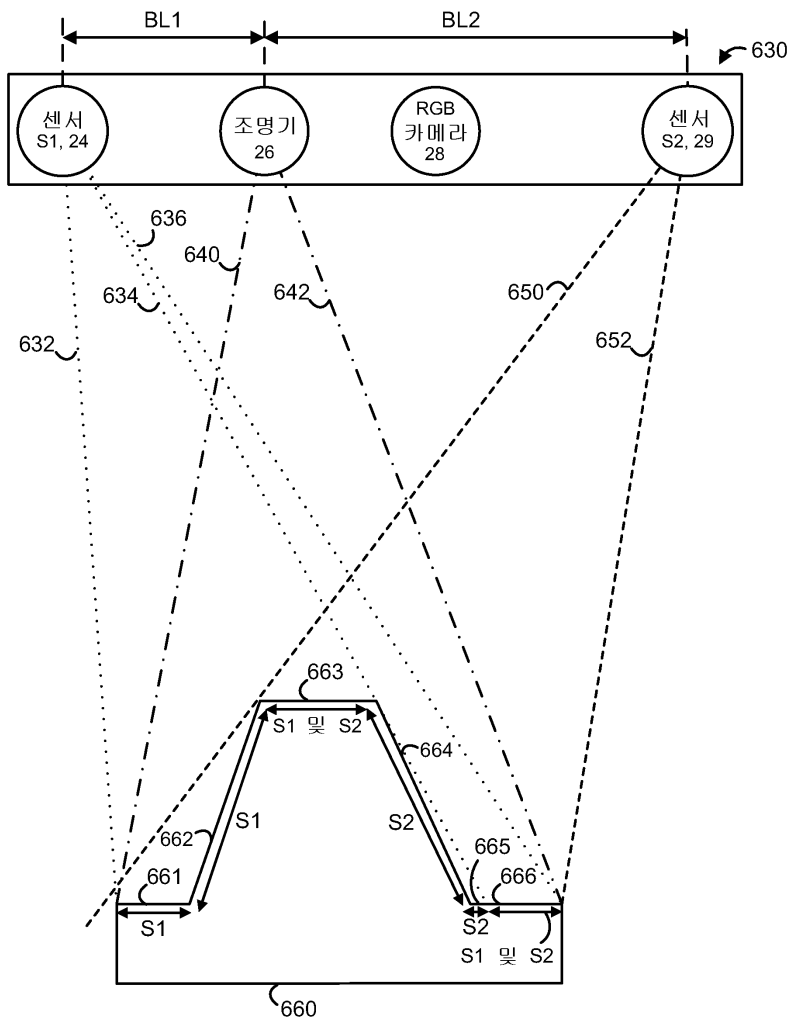
도면6b



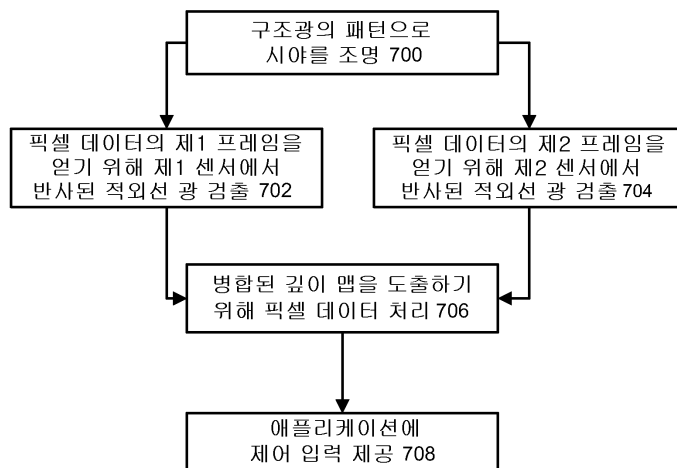
도면6c



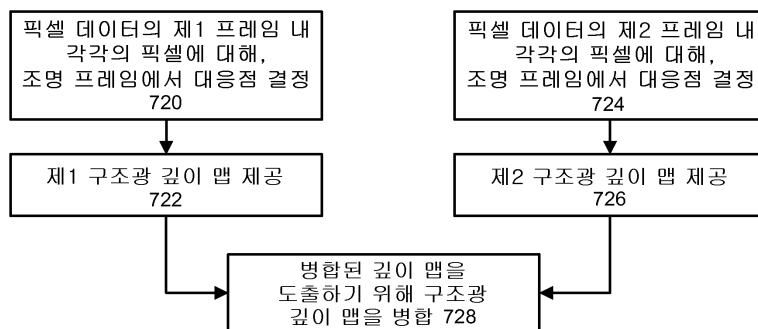
도면6d



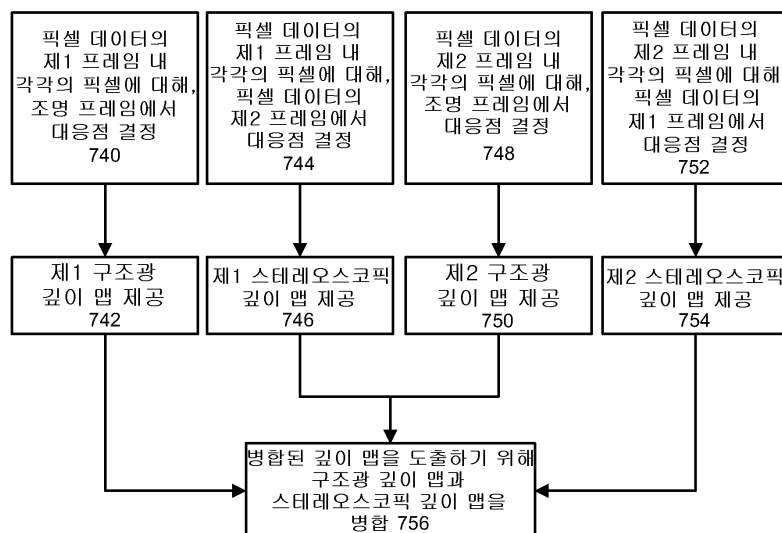
도면7a



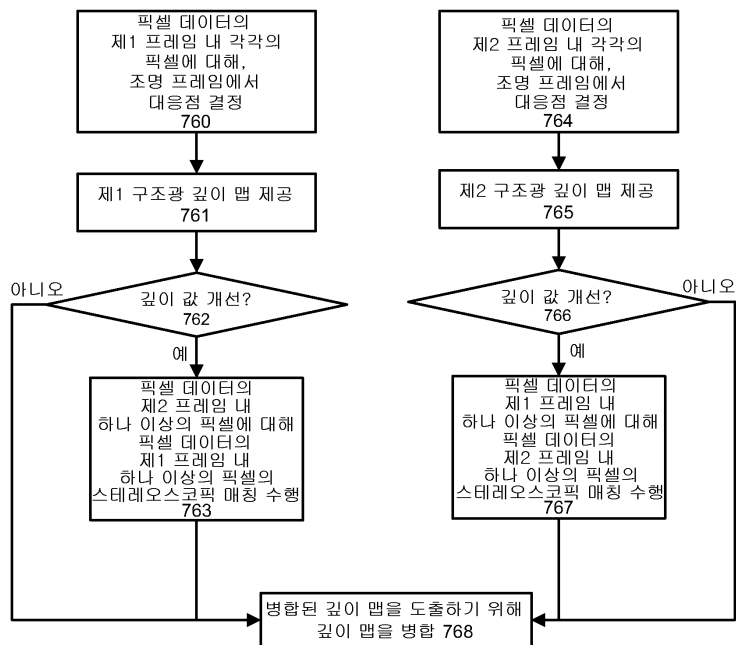
도면7b



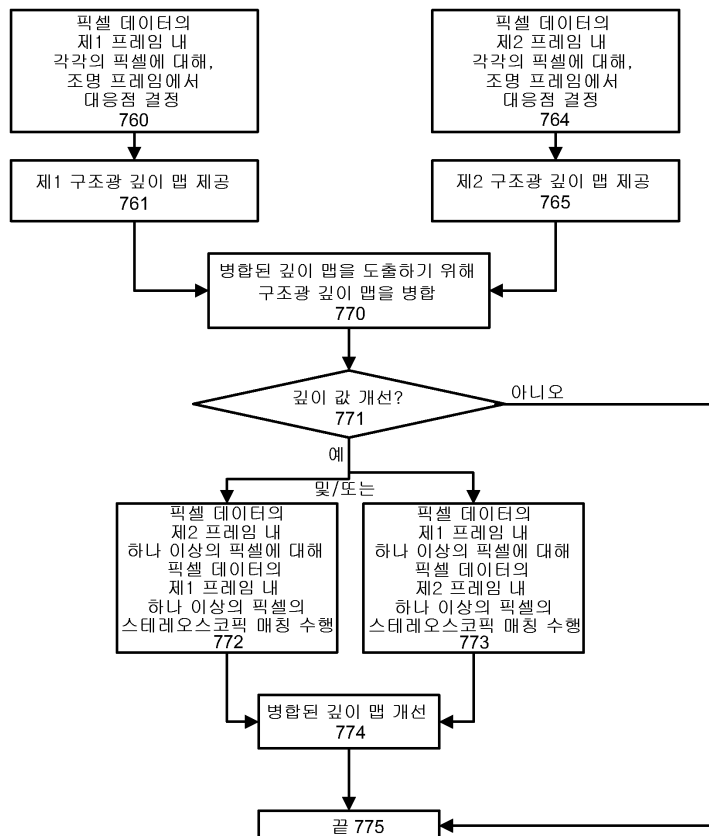
도면7c



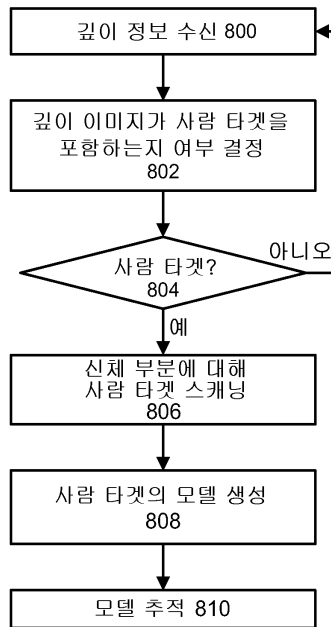
도면7d



도면7e



도면8



도면9

