(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2017년08월31일
(11) 등록번호 10-1772890
(24) 등록일자 2017년08월24일

(51) 국제특허분류(Int. Cl.)
G06K 9/00 (2006.01) *G06K 9/62* (2006.01)
(52) CPC특허분류
G06K 9/00355 (2013.01)
G06F 3/017 (2013.01)
(21) 출원번호 10-2016-0037120
(22) 출원일자 2016년03월28일
심사청구일자 2016년03월28일
(56) 선행기술조사문헌
JP05802667 B2
JP2012079177 A
KR1020140007367 A

(73) 특허권자
한국과학기술원
대전광역시 유성구 대학로 291(구성동)
(72) 발명자
우운택
대전광역시 유성구 대학로 291 (구성동)
장영균
대전광역시 유성구 대학로 291 (구성동)
(뒷면에 계속)
(74) 대리인
특허법인충정

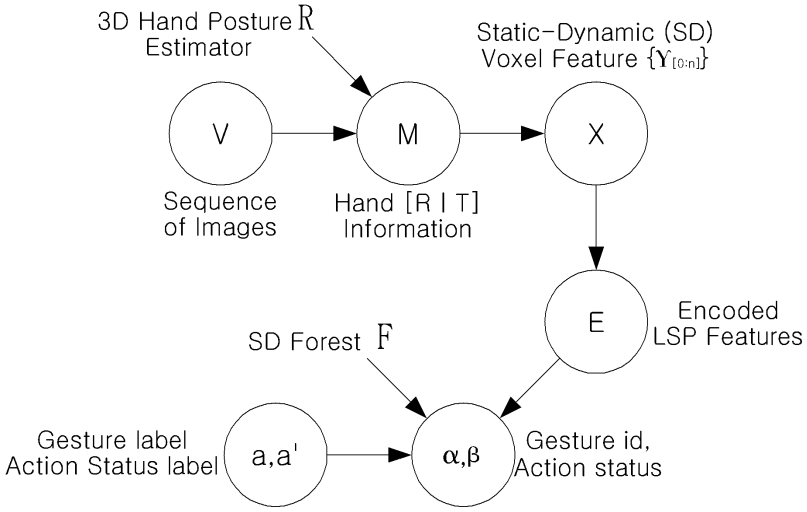
전체 청구항 수 : 총 14 항 심사관 : 강현일

(54) 발명의 명칭 **손의 홀딩 제스처에 대응하는 물체의 종류 및 조작 제스처에 대응하는 조작의 종류를 판단하기 위한 제스처 감지 장치 및 이에 의한 제스처 감지 방법**

(57) 요 약

사용자에게 착용된 깊이 카메라에 의해 촬영되는 손의 제 1 시퀀스 영상을 획득하는 단계; 획득한 제 1 시퀀스 영상 내 복수의 프레임 각각의 손에 복수의 복셀(voxel)로 나뉘어진 사각 큐브(rectangular cube)를 매핑하는 단계; 복수의 프레임 각각에 대해, 사각 큐브를 구성하는 복수의 복셀들의 정적 복셀 특징(static voxel feature)을 추출하고, 추출된 정적 복셀 특징을 인코딩하여 제 1 LSP 특징(layer shape pattern feature)을 획득하는 단계; 및 제 1 LSP 특징을 랜덤 포레스트에 입력하여, 제 1 시퀀스 영상에서 손의 홀딩 제스처에 대응하는 물체의 종류와 손의 조작 제스처에 대응하는 조작의 종류를 판단하는 단계를 포함하는 것을 특징으로 하는 본 발명의 일 실시예에 따른 제스처 감지 장치에 의한 제스처 감지 방법이 개시된다.

대 표 도 - 도1



(52) CPC특허분류
G06K 9/6202 (2013.01)
(72) 발명자
전익범
대전광역시 유성구 대학로 291 (구성동)
김태균
대전광역시 유성구 대학로 291 (구성동)

이 발명을 지원한 국가연구개발사업
과제고유번호 2014R1A2A2A01003005
부처명 미래창조과학부
연구관리전문기관 한국연구재단
연구사업명 이공분야기초연구사업 - 중견연구자지원사업
연구과제명 3D 글래스 환경에서 기하관계 중심의 스마트폰 기반 상호작용형 저작 기술
기 여 율 1/1
주관기관 한국과학기술원
연구기간 2014.05.01 ~ 2017.04.30

명 세 서

청구범위

청구항 1

사용자에게 착용된 깊이 카메라에 의해 촬영되는 손의 제 1 시퀀스 영상을 획득하는 단계;

상기 획득한 제 1 시퀀스 영상 내 복수의 프레임 각각의 손에 복수의 복셀(voxel)로 나뉘어진 사각 큐브(rectangular cube)를 매핑하는 단계;

상기 복수의 프레임 각각에 대해, 상기 사각 큐브를 구성하는 복수의 복셀들의 정적 복셀 특징(static voxel feature)을 추출하고, 추출된 정적 복셀 특징을 인코딩하여 제 1 LSP 특징(layer shape pattern feature)을 획득하는 단계; 및

상기 제 1 LSP 특징을 랜덤 포레스트(random forest)에 입력하여, 상기 제 1 시퀀스 영상에서 손의 홀딩(holding) 제스처에 대응하는 물체의 종류와 손의 조작(manipulation) 제스처에 대응하는 조작의 종류를 판단하는 단계를 포함하고,

상기 랜덤 포레스트는, 홀딩 제스처에 대응하는 물체의 종류 및 조작 제스처에 대응하는 조작의 종류를 알고 있는 손의 제 2 시퀀스 영상의 복수의 프레임 각각으로부터 획득된 제 2 LSP 특징에 기초하여 구성된 것을 특징으로 하는, 제스처 감지 장치에 의한 제스처 감지 방법.

청구항 2

제1항에 있어서,

상기 사각 큐브를 매핑하는 단계는,

상기 복수의 프레임 각각의 손을 구성하는 복수의 3차원 픽셀들을 내부에 포함하도록 상기 사각 큐브를 손에 매핑하는 단계를 포함하는 것을 특징으로 하는 제스처 감지 방법.

청구항 3

제1항에 있어서,

상기 정적 복셀 특징은,

상기 복수의 복셀들 각각의 내부에 포함된 상기 손의 3차원 픽셀들의 개수를 엘리먼트로 포함하는 것을 특징으로 하는 제스처 감지 방법.

청구항 4

제3항에 있어서,

상기 제스처 감지 방법은,

상기 사각 큐브를 구성하는 복수의 복셀 각각의 내부에 포함된 상기 손의 3차원 픽셀들의 개수를 정규화(normalization)하는 단계를 더 포함하는 것을 특징으로 하는 제스처 감지 방법.

청구항 5

제1항에 있어서,

상기 제 1 LSP 특징은,

상기 사각 큐브를 구성하는 복수의 복셀 플레인(voxel plane) 각각의 내부에 3차원 픽셀들이 포함된 정도를 나타내는 것을 특징으로 하는 제스처 감지 방법.

청구항 6

제5항에 있어서,

상기 제 1 LSP 특징을 획득하는 단계는,

상기 획득한 제 1 시퀀스 영상 내 복수의 프레임 각각에서 손의 손바닥의 지향(orientation) 방향을 결정하는 단계;

상기 제 1 시퀀스 영상 내 복수의 프레임 각각에서의 소정 지점을 원점으로 하고, 상기 결정된 지향 방향에 따라 로컬 좌표의 각 축(axis)의 방향을 결정하는 단계; 및

상기 복수의 복셀 플레인에 대해 상기 로컬 좌표 상에서의 인덱스를 매핑하는 단계를 포함하는 것을 특징으로 하는 제스처 감지 방법.

청구항 7

제5항에 있어서,

로컬 좌표의 제 1 축 상의 어느 하나의 복셀 플레인에 대한 LSP 엘리먼트는,

상기 어느 하나의 복셀 플레인이 제 2 축 또는 제 3 축을 따라 n개의 층으로 구분되어 있는 경우, 각 층의 복셀 내의 3차원 픽셀의 밀도에 대응하여 상기 n개의 층 각각에 할당된 값이 나열된 값을 포함하는 것을 특징으로 하는 제스처 감지 방법.

청구항 8

제7항에 있어서,

상기 n개의 층 각각에 할당된 값을 나열할 때, 원점으로부터 가장 멀리 떨어진 층에 할당된 값은 가장 큰 자릿수(positional number)를 갖고, 상기 원점과 가장 가까운 층에 할당된 값은 가장 작은 자릿수를 갖도록 나열되는 것을 특징으로 하는 제스처 감지 방법.

청구항 9

제1항에 있어서,

상기 제스처 감지 방법은,

상기 랜덤 포레스트를 구성하는 단계를 더 포함하고,

상기 랜덤 포레스트를 구성하는 단계는,

상기 랜덤 포레스트의 트리 내 최상위 층의 스플릿 노드(split node)의 분할 함수(split function)를 정적(static) 함수로 결정하고, 상기 최상위 층의 스플릿 노드의 분할 함수의 파라미터를 최적화하는 단계; 및

상기 최상위 층의 스플릿 노드에서 나뉘어진 학습 데이터 세트를 입력받는 하위층의 스플릿 노드에 대해서는, 상기 학습 데이터 세트 중 동일한 홀딩 제스처 라벨로 라벨링된 데이터의 가장 큰 확률 두 번째로 큰 확률 사이의 차이가 임계값보다 큰 경우, 분할 함수를 동적(dynamic) 함수로 결정하고, 상기 차이가 임계값 이하인 경우, 분할 함수를 정적 함수로 결정하는 단계를 포함하는 것을 특징으로 하는 제스처 감지 방법.

청구항 10

제9항에 있어서,

상기 랜덤 포레스트를 구성하는 단계는,

상기 하위층의 스플릿 노드의 분할 함수가 동적 함수로 결정된 경우, 상기 하위층의 스플릿 노드의 하위층에 위치하는 모든 스플릿 노드에 대해서는 분할 함수를 동적 함수로 결정하는 단계를 포함하는 것을 특징으로 하는 제스쳐 감지 방법.

청구항 11

제9항에 있어서,

상기 정적 함수는, 어느 하나의 프레임에서 획득된 LSP 특징에 포함된 LSP 엘리먼트를 나타내고,

상기 동적 함수는, 어느 하나의 프레임에서 획득된 LSP 특징에 포함된 LSP 엘리먼트와 다른 하나의 프레임에서 획득된 LSP 특징에 포함된 LSP 엘리먼트의 차이를 나타내는 것을 특징으로 하는 제스쳐 감지 방법.

청구항 12

제1항에 있어서,

상기 판단하는 단계는,

상기 랜덤 포레스트의 복수의 트리 각각에서 결정된 물체의 종류와 조작의 종류의 확률의 평균과 기 설정된 임계 값을 비교하여, 상기 홀딩 제스쳐에 대응하는 물체의 종류 및 조작 제스쳐에 대응하는 조작의 종류를 판단하는 단계를 포함하는 것을 특징으로 하는 제스쳐 감지 방법.

청구항 13

하드웨어와 결합되어 제1항 내지 제12항 중 어느 하나의 항의 제스쳐 감지 방법을 실행시키기 위하여 매체에 저장된 컴퓨터 프로그램.

청구항 14

사용자에게 착용된 깊이 카메라에 의해 촬영되는 손의 제 1 시퀀스 영상을 획득하는 영상 획득부;

상기 획득한 제 1 시퀀스 영상 내 복수의 프레임 각각의 손에 복수의 복셀로 나뉘어진 사각 큐브를 매핑하는 영상 처리부;

상기 복수의 프레임 각각에 대해, 상기 사각 큐브를 구성하는 복수의 복셀들의 정적 복셀 특징을 추출하고, 추출된 정적 복셀 특징을 인코딩하여 제 1 LSP 특징을 획득하는 특징 추출부; 및

상기 제 1 LSP 특징을 랜덤 포레스트에 입력하여, 상기 제 1 시퀀스 영상에서 손의 홀딩 제스쳐에 대응하는 물체의 종류와 해당 물체에 대한 조작의 종류를 판단하는 판단부를 포함하고,

상기 랜덤 포레스트는, 홀딩 제스쳐에 대응하는 물체의 종류 및 조작 제스쳐에 대응하는 조작의 종류를 알고 있는 손의 제 2 시퀀스 영상의 복수의 프레임 각각으로부터 획득된 제 2 LSP 특징에 기초하여 구성된 것을 특징으로 하는 제스쳐 감지 장치.

발명의 설명

기 술 분 야

[0001] 본 발명은 가상 현실 및 증강 현실 분야에 관한 것이다. 보다 구체적으로, 본 발명은 손의 홀딩 제스처에 대응하는 물체의 종류 및 조작 제스처에 대응하는 조작의 종류를 판단하기 위한 제스처 감지 장치 및 이에 의한 제스처 감지 방법에 관한 것이다.

배 경 기 술

[0002] 최근의 HMD(head mounted display)의 발전은 증강 현실(augmented reality: AR) 및 가상 현실(virtual reality: VR) 환경에서 가상 물체와 상호작용을 하는 것과 같은 다양한 경험을 제공한다. HMD를 착용한 상태에서 가상 물체와 상호작용을 하는데 있어 맨 손(bare hand)를 사용하는 것이 가장 자연스러운 방법이다. AR 및 VR에서 맨손 기반의 제스처 감지에 대한 연구들은 예를 들어, 장면 네비게이션(scene navigation)(참고 문헌 [1]), 물체 이동(참고 문헌 [2]), 3D 모델 어셈블리(참고 문헌 [3]) 등과 같은 상호 작용 시나리오를 제안하고 있다. 그러나, 이들 맨손 기반의 상호 작용의 대부분은 일인칭 시점(egocentric viewpoint)에서의 상호 작용에 활용하기는 어렵다. 이는, 일인칭 시점에서는 손의 일부분에 의해 가상 정보가 자주 사라지기(occluded) 때문이다. 더 나아가, 매우 작은(손톱 크기의) 물체의 선택 및 회전 조작과 같은 정교한 상호 작용은 여전히 어려운 문제점으로 남아있다.

[0004] 본 명세서에서 참조되는 참고 문헌의 리스트는 다음과 같다.

[0006] [1] Z. Pan, Y. Li, M. Zhang, C. Sun, K. Guo, X. Tang, and S. Z. Zhou. A real-time multi-cue hand tracking algorithm based on computer vision. In Proceedings of the 2010 IEEE Virtual Reality Conference, VR '10, pages 219-222, Washington, DC, USA, 2010. IEEE Computer Society.

[0007] [2] G. Hackenberg, R. McCall, and W. Broll. Lightweight palm and finger tracking for real-time 3D gesture control. In Virtual Reality Conference (VR), 2011 IEEE, pages 19 -26, march 2011.

[0008] [3] R. Wang, S. Paris, and J. Popovi'c. 6D hands: Markerless hand-tracking for computer aided design. In Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology, UIST '11, pages 549-558, New York, NY, USA, 2011. ACM.

[0009] [4] S. Melax, L. Keselman, and S. Orsten. Dynamics based 3D skeletal hand tracking. In Proceedings of the 2013 Graphics Interface Conference, GI '13, pages 63-70, Toronto, Ont., Canada, Canada, 2013.

[0010] [5] D. Tang, T.-H. Yu, and T.-K. Kim. Real-time articulated hand pose estimation using semi-supervised transductive regression forests. In The IEEE International Conference on Computer Vision (ICCV), December 2013.

[0011] [6] H. Liang and J. Yuan. Hand parsing and gesture recognition with a commodity depth camera. In Computer Vision and Machine Learning with RGB-D Sensors, pages 239-265. Springer, 2014.

[0012] [7] H. Liang, J. Yuan, and D. Thalmann. Parsing the hand in depth images. IEEE Trans. on Multimedia, 16(5):1241-1253, 2014.

[0013] [8] Y. Jang, S.-T. Noh, H. J. Chang, T.-K. Kim, and W. Woo. 3D Finger CAPE: Clicking action and position estimation under self-occlusions in egocentric viewpoint. IEEE Trans. on Vis. Comput. Graph., 21(4):0-10, April 2015.

발명의 내용

해결하려는 과제

[0015] 본 발명의 일 실시예에 따른 제스처 감지 장치 및 이에 의한 제스처 감지 방법은 사용자에게 보다 정교한 AR 및 VR 상호 작용을 제공하는 것을 목적으로 한다.

[0016] 또한, 본 발명의 일 실시예에 따른 제스처 감지 장치 및 이에 의한 제스처 감지 방법은 사용자에게 착용되는 깊이 카메라 외에 사용자에게 별도의 장치의 장착을 요구하지 않는 것을 목적으로 한다.

[0017] 또한, 본 발명의 일 실시예에 따른 제스처 감지 장치 및 이에 의한 제스처 감지 방법은 AR 및 VR 환경에서 사용

자의 홀딩 제스처에 기초하여 사용자가 조작하고자 하는 물체를 정확하게 소환(summon)하고, 소환된 물체에 대한 자유로운 조작을 가능하게 하는 것을 목적으로 한다.

과제의 해결 수단

[0018] 본 발명의 일 실시예에 따른 제스처 감지 장치에 의한 제스처 감지 방법은,

[0019] 사용자에게 착용된 깊이 카메라에 의해 촬영되는 손의 제 1 시퀀스 영상을 획득하는 단계; 상기 획득한 제 1 시퀀스 영상 내 복수의 프레임 각각의 손에 복수의 복셀(voxel)로 나뉘어진 사각 큐브(rectangular cube)를 매핑하는 단계; 상기 복수의 프레임 각각에 대해, 상기 사각 큐브를 구성하는 복수의 복셀들의 정적 복셀 특징(static voxel feature)을 추출하고, 추출된 정적 복셀 특징을 인코딩하여 제 1 LSP 특징(layer shape pattern feature)을 획득하는 단계; 및 상기 제 1 LSP 특징을 랜덤 포레스트(random forest)에 입력하여, 상기 제 1 시퀀스 영상에서 손의 홀딩(holding) 제스처에 대응하는 물체의 종류와 손의 조작(manipulation) 제스처에 대응하는 조작의 종류를 판단하는 단계를 포함하고, 상기 랜덤 포레스트는, 홀딩 제스처에 대응하는 물체의 종류 및 조작 제스처에 대응하는 조작의 종류를 알고 있는 손의 제 2 시퀀스 영상의 복수의 프레임 각각으로부터 획득된 제 2 LSP 특징에 기초하여 구성될 수 있다.

[0020] 상기 사각 큐브를 매핑하는 단계는, 상기 복수의 프레임 각각의 손을 구성하는 복수의 3차원 픽셀들을 내부에 포함하도록 상기 사각 큐브를 손에 매핑하는 단계를 포함할 수 있다.

[0021] 상기 정적 복셀 특징은, 상기 복수의 복셀들 각각의 내부에 포함된 상기 손의 3차원 픽셀들의 개수를 엘리먼트로 포함할 수 있다.

[0022] 상기 제스처 감지 방법은, 상기 사각 큐브를 구성하는 복수의 복셀 각각의 내부에 포함된 상기 손의 3차원 픽셀들의 개수를 정규화(normalization)하는 단계를 더 포함할 수 있다.

[0023] 상기 제 1 LSP 특징은, 상기 사각 큐브를 구성하는 복수의 복셀 플레인(voxel plane) 각각의 내부에 상기 3차원 픽셀들이 포함된 정도를 나타낼 수 있다.

[0024] 상기 제 1 LSP 특징을 획득하는 단계는, 상기 획득한 제 1 시퀀스 영상 내 복수의 프레임 각각에서 손의 손바닥의 지향(orientation) 방향을 결정하는 단계; 상기 제 1 시퀀스 영상 내 복수의 프레임 각각에서의 소정 지점을 원점으로 하고, 상기 결정된 지향 방향에 따라 로컬 좌표의 각 축(axis)의 방향을 결정하는 단계; 및 상기 복수의 복셀 플레인에 대해 상기 로컬 좌표 상에서의 인덱스를 매핑하는 단계를 포함할 수 있다.

[0025] 로컬 좌표의 제 1 축 상의 어느 하나의 복셀 플레인에 대한 LSP 엘리먼트는, 상기 어느 하나의 복셀 플레인이 제 2 축 또는 제 3 축을 따라 n개의 층으로 구분되어 있는 경우, 각 층의 복셀 내의 3차원 픽셀의 밀도에 대응하여 상기 n개의 층 각각에 할당된 값이 나열된 값을 포함할 수 있다.

[0026] 상기 n개의 층 각각에 할당된 값을 나열할 때, 원점으로부터 가장 멀리 떨어진 층에 할당된 값은 가장 큰 자릿수(positional number)를 갖고, 상기 원점과 가장 가까운 층에 할당된 값은 가장 작은 자릿수를 갖도록 나열될 수 있다.

[0027] 상기 제스처 감지 방법은, 상기 랜덤 포레스트를 구성하는 단계를 더 포함하고, 상기 랜덤 포레스트를 구성하는 단계는, 상기 랜덤 포레스트의 트리 내 최상위 층의 스플릿 노드(split node)의 분할 함수(split function)를 정적(static) 함수로 결정하고, 상기 최상위 층의 스플릿 노드의 분할 함수의 파라미터를 최적화하는 단계; 및 상기 최상위 층의 스플릿 노드에서 나뉘어진 학습 데이터 세트를 입력받는 하위층의 스플릿 노드에 대해서는, 상기 학습 데이터 세트 중 동일한 홀딩 제스처 라벨로 라벨링된 데이터의 가장 큰 확률 두 번째로 큰 확률 사이의 차이가 임계값보다 큰 경우, 분할 함수를 동적(dynamic) 함수로 결정하고, 상기 차이가 임계값 이하인 경우, 분할 함수를 정적 함수로 결정하는 단계를 포함할 수 있다.

[0028] 상기 랜덤 포레스트를 구성하는 단계는, 상기 하위층의 스플릿 노드의 분할 함수가 동적 함수로 결정된 경우, 상기 하위층의 스플릿 노드의 하위층에 위치하는 모든 스플릿 노드에 대해서는 분할 함수를 동적 함수로 결정하는 단계를 포함할 수 있다.

[0029] 상기 정적 함수는, 어느 하나의 프레임에서 획득된 LSP 특징에 포함된 LSP 엘리먼트를 나타내고, 상기 동적 함수는, 어느 하나의 프레임에서 획득된 LSP 특징에 포함된 LSP 엘리먼트와 다른 하나의 프레임에서 획득된 LSP 특징에 포함된 LSP 엘리먼트의 차이를 나타낼 수 있다.

[0030] 상기 판단하는 단계는, 상기 랜덤 포레스트의 복수의 트리 각각에서 결정된 물체의 종류와 조작의 종류의 확률

의 평균과 기 설정된 임계 값을 비교하여, 상기 홀딩 제스처에 대응하는 물체의 종류 및 조작 제스처에 대응하는 조작의 종류를 판단하는 단계를 포함할 수 있다.

[0031] 본 발명의 다른 실시예에 따른 제스처 감지 장치는, 사용자에게 착용된 깊이 카메라에 의해 촬영되는 손의 제 1 시퀀스 영상을 획득하는 영상 획득부; 상기 획득한 제 1 시퀀스 영상 내 복수의 프레임 각각의 손에 복수의 복셀로 나뉘어진 사각 큐브를 매핑하는 영상 처리부; 상기 복수의 프레임 각각에 대해, 상기 사각 큐브를 구성하는 복수의 복셀들의 정적 복셀 특징을 추출하고, 추출된 정적 복셀 특징을 인코딩하여 제 1 LSP 특징을 획득하는 특징 추출부; 및 상기 제 1 LSP 특징을 랜덤 포레스트에 입력하여, 상기 제 1 시퀀스 영상에서 손의 홀딩 제스처에 대응하는 물체의 종류와 해당 물체에 대한 조작의 종류를 판단하는 판단부를 포함하고, 상기 랜덤 포레스트는, 홀딩 제스처에 대응하는 물체의 종류 및 조작 제스처에 대응하는 조작의 종류를 알고 있는 손의 제 2 시퀀스 영상의 복수의 프레임 각각으로부터 획득된 제 2 LSP 특징에 기초하여 구성될 수 있다.

발명의 효과

[0033] 본 발명의 일 실시예에 따른 제스처 감지 장치 및 이에 의한 제스처 감지 방법이 달성할 수 있는 일부의 효과는 다음과 같다.

[0035] i) 사용자에게 보다 정교한 AR 및 VR 상호 작용을 제공할 수 있다.

[0036] ii) 사용자에게 착용되는 깊이 카메라 외에 사용자에게 별도의 장치의 장착을 요구하지 않는다.

[0037] iii) AR 및 VR 환경에서 사용자의 홀딩 제스처에 기초하여 사용자가 조작하고자 하는 물체를 정확하게 소환하고, 소환된 물체에 대한 자유로운 조작을 가능하게 한다.

[0039] 다만, 본 발명의 일 실시예에 따른 제스처 감지 장치 및 이에 의한 제스처 감지 방법이 달성할 수 있는 효과는 이상에서 언급한 것들로 제한되지 않으며, 언급하지 않은 또 다른 효과들은 아래의 기재로부터 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

도면의 간단한 설명

[0040] 도 1은 본 발명의 일 실시예에 따른 제스처 감지 방법을 설명하기 위한 개략적인 도면이다.

도 2는 손의 글로벌 좌표를 로컬 좌표로 변환하는 내용을 설명하기 위한 도면이다.

도 3은 사각 큐브로부터 추출된 정적 복셀 특징을 설명하기 위한 도면이다.

도 4 및 도 5는 정적 복셀 특징을 LSP 특징으로 인코딩하는 방법을 설명하기 위한 도면이다.

도 6은 손의 제 2 시퀀스 영상의 복수의 프레임 각각으로부터 획득된 제 2 LSP 특징에 기초하여 구성된 랜덤 포레스트의 트리를 나타내는 예시적인 도면이다.

도 7(a) 및 도 7(b)는 종래의 기술과 본 발명의 일 실시예에 따라 판단되는 물체의 종류의 정확도를 비교하기 위한 도면이다.

도 8은 종래의 기술과 본 발명의 일 실시예에 따라 판단되는 조작의 종류의 정확도를 비교하기 위한 도면이다.

도 9는 본 발명의 다른 실시예에 따른 제스처 감지 장치의 구성을 나타내는 블록도이다.

발명을 실시하기 위한 구체적인 내용

[0041] 본 발명은 다양한 변경을 가할 수 있고 여러 가지 실시예를 가질 수 있는 바, 특정 실시예들을 도면에 예시하고, 이를 상세한 설명을 통해 상세히 설명하고자 한다. 그러나, 이는 본 발명을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 본 발명은 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다.

[0042] 본 발명을 설명함에 있어서, 관련된 공지 기술에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우 그 상세한 설명을 생략한다. 또한, 본 명세서의 설명 과정에서 이용되는 숫자(예를 들어, 제 1, 제 2 등)는 하나의 구성요소를 다른 구성요소와 구분하기 위한 식별기호에 불과하다.

[0043] 또한, 본 명세서에서, 일 구성요소가 다른 구성요소와 "연결된다" 거나 "접속된다" 등으로 언급된 때에는, 상기 일 구성요소가 상기 다른 구성요소와 직접 연결되거나 또는 직접 접속될 수도 있지만, 특별히 반대되는 기재가

존재하지 않는 이상, 중간에 또 다른 구성요소를 매개하여 연결되거나 또는 접속될 수도 있다고 이해되어야 할 것이다.

[0044] 또한, 본 명세서에서 '~부(유닛)', '모듈' 등으로 표현되는 구성요소는 소프트웨어, FPGA 또는 ASIC과 같은 하드웨어 구성요소를 의미하며, 이 구성요소는 어떤 역할들을 수행한다. 그렇지만, 구성 요소는 소프트웨어 또는 하드웨어에 한정되는 의미는 아니다. 구성요소는 어드레싱할 수 있는 저장 매체에 있도록 구성될 수도 있다. 또한, 2개 이상의 구성요소가 하나의 구성요소로 합쳐지거나 또는 하나의 구성요소가 보다 세분화된 기능별로 2개 이상으로 분화될 수도 있다. 또한, 이하에서 설명할 구성요소 각각은 자신이 담당하는 주기능 이외에도 다른 구성요소가 담당하는 기능 중 일부 또는 전부의 기능을 추가적으로 수행할 수도 있으며, 구성요소 각각이 담당하는 주기능 중 일부 기능이 다른 구성요소에 의해 전담되어 수행될 수도 있음은 물론이다.

[0045] 본 발명은 사용자의 요구에 따라 사용자의 손에 물체(예를 들어, 스마트폰)를 소환(summon)하고, 소환된 물체를 조작할 수 있게 한다는 개념으로부터 도출되었다. 일반적으로, 사용자는 일상적으로 많이 사용되는 장비의 홀딩 제스쳐나 이를 조작하는 조작 제스쳐에 대해 학습하고 있다. 이러한 제스쳐는 직관적이며, 기억하기 쉽다. 본 발명의 일 실시예에서는 사용자가 실제의 물체를 홀딩하고 있지 않더라도 사용자의 손의 제스쳐만을 가지고 사용자가 어떠한 물체를 홀딩하고, 어떠한 방식으로 조작하고자 하는지를 판단할 수 있다.

[0046] 도 1은 본 발명의 일 실시예에 따른 제스쳐 감지 방법을 설명하기 위한 개략적인 도면이다.

[0047] 본 발명의 일 실시예에 따른 제스쳐 감지 장치는 사용자에게 착용된 깊이 카메라(depth camera)에 의해 촬영되는 손의 제 1 시퀀스 영상(V)을 획득한다. 제 1 시퀀스 영상(V)은 복수의 프레임을 포함할 수 있다. 깊이 카메라는 사용자에게 착용되는 HMD에 장착될 수 있다.

[0048] 제스쳐 감지 장치는 일반적인 3D 손바닥 자세 추정기(palm pose estimator) $\mathscr{R}$ 로 손바닥의 회전 R 및 이동 T를 나타내는 변환 정보 M을 추정한다. 여기서, 3D 손바닥 자세 추정기는 예를 들어, 참고 문헌 [4]가 참조될 수 있다. 그리고, 제스쳐 감지 장치는 손의 회전이나 이동에 영향을 받지 않는 특징을 추출하기 위해 손의 3차원 픽셀의 좌표를 변환한다.

[0049] 도 2는 손의 글로벌 좌표를 로컬 좌표로 변환하는 내용을 설명하기 위한 도면으로서, 도면부호 210은 인덱스 0에 대응하는 프레임에 포함된 손, 도면부호 220은 인덱스 k에 대응하는 프레임에 포함된 손을 나타내며, 동일한 손의 새끼손가락이라도 카메라 중심의 카메라 좌표(camera coordinate)에 따르면 서로 상이한 좌표를 갖게 된다. 제스쳐 감지 장치는 손의 3차원 픽셀들의 카메라 좌표를 아래의 수학식 1에 따라 각 프레임의 소정 지점(p)을 원점으로 하는 로컬 좌표(local coordinate) 또는 손 좌표(hand coordinate)로 변환할 수 있다.

[0051] [수학식 1]

$$v_l = M_0 v_{g0} = M_k v_{gk} = \begin{pmatrix} R_k & T_k \\ 0^\top & 1 \end{pmatrix} v_{gk}$$

[0052]

[0053] 상기 수학식 1에서 v_l 은 인덱스 0 및 인덱스 k에 대응하는 프레임에 포함된 손의 3차원 픽셀의 로컬 좌표, v_{g0} 및 v_{gk} 각각은 인덱스 0 및 인덱스 k에 대응하는 프레임에 포함된 손의 3차원 픽셀의 카메라 좌표, M_0 및 M_k 각각은 인덱스 0 및 인덱스 k에 대응하는 프레임에 포함된 손의 변환 정보로서, 3 by 3의 회전 매트릭스(rotation matrix) R과 3 by 1의 이동 매트릭스(translation matrix) T로 구성된 손바닥 변환 매트릭스를 나타낸다.

[0054] 다시 도 1을 참조하면, 제스쳐 감지 장치는 각 프레임의 정적 복셀 특징 x_i(i는 프레임의 인덱스로서, 0 내지 m)의 집합인 정적-동적 복셀 특징 $X = x_{[0:m]}$ (m은 복수의 프레임 중 마지막 프레임의 인덱스)을 추출한다. 정적 복셀 특징 x_i는 복수의 복셀들($\Upsilon_i = (C_z, N_c)$)을 엘리먼트로 포함하며, Cz는 복셀 Υ_i 내에 포함된 3차원 픽

셀의 평균 깊이 값, Nc는 복셀 Υ_i 내에 포함된 3차원 픽셀의 개수를 나타낸다.

[0055] 도 3은 사각 큐브(300)로부터 추출된 정적 복셀 특징을 설명하기 위한 도면으로서, 기 설정된 개수의 복셀로 나뉘어진 사각 큐브(300)가 손의 3차원 픽셀들을 포함하도록 손에 매핑된 경우, 각 복셀에 포함된 3차원 픽셀의 개수들이 정적 복셀 특징 x_i로 포함될 수 있다. 복수의 프레임 중 첫 번째 프레임에 대응하는 정적 복셀 특징 x_0로부터 마지막 프레임에 대응하는 정적 복셀 특징 x_m의 집합은 제 1 시퀀스 영상의 정적-동적 복셀 특징 X에 해당한다.

[0056] 제스쳐 감지 장치는 복셀에 포함된 3차원 픽셀의 개수가 획득되면, 거리에 영향을 받지 않도록 3차원 픽셀의 개수를 정규화할 수 있다. 만약, 손이 카메라에 더 가까워진다면, 하나의 복셀에 더 많은 3차원 픽셀이 포함되게 되기 때문이다. 제스쳐 감지 장치는 i번째 복셀인 Υ_i 내의 3차원 픽셀의 개수인 $\Upsilon_i(\mathrm{N}_c)$는 다음의 수학식 2에 따라 정규화된다.

[0058] [수학식 2]

$$\Upsilon_i(\hat{\mathrm{N}}_c) = b + \frac{c}{\Upsilon_i(\mathrm{C}_z)} + \Upsilon_i(\mathrm{N}_c)$$

[0059]

[0060] 수학식 2에서 $\Upsilon_i(\hat{\mathrm{N}}_c)$는 복셀 Υ_i 내의 정규화된 3차원 픽셀의 개수, $\Upsilon_i(\mathrm{N}_c)$는 정규화되기 전의 복셀 Υ_i 내의 3차원 픽셀의 개수, $\Upsilon_i(\mathrm{C}_z)$는 복셀 Υ_i 내의 3차원 픽셀의 카메라 좌표에서의 평균 깊이 값, b는 정규화된 3차원 픽셀의 개수가 음수(negative number)가 되지 않도록 3차원 픽셀의 개수를 이동시키기 위한 상수 값을 나타낸다.

[0061] 또한, 상기 수학식 2에서 c는 3차원 픽셀 개수를 정규화시키기 위해 적용되는 상수 인자 값을 나타내며, 아래의 수학식 3에 따라 계산될 수 있다.

[0063] [수학식 3]

$$c = \left(\Upsilon_i(\mathrm{N}_c)_2 - \Upsilon_i(\mathrm{N}_c)_1\right) \times \frac{\Upsilon_i(\mathrm{C}_z)_1 \times \Upsilon_i(\mathrm{C}_z)_2}{\Upsilon_i(\mathrm{C}_z)_2 - \Upsilon_i(\mathrm{C}_z)_1}$$

[0064]

[0065] 수학식 3에서 $\Upsilon_i(\cdot)_1$, $\Upsilon_i(\cdot)_2$ 각각은 두 개의 서로 다른 프레임 내에서 캡쳐(capture)된 동일한 i번째 복셀 Υ_i를 나타낸다. 제스쳐 감지 장치는 c 값을 획득하기 위해 가장 큰 깊이 차이를 갖는 두 개의 프레임을 의도적으로 선택할 수 있다.

[0066] 도 3에 도시된 바와 같이, 하나의 사각 큐브에 총 75개의 복셀이 포함된 경우, 정적 복셀 특징 x_i는 총 75개의 엘리먼트를 갖게 된다. 모든 복셀에 대한 3차원 픽셀 개수를 엘리먼트로 포함하는 경우, 그 개수가 매우 많아지게 되므로, 본 발명의 일 실시예에서는 도 1에 도시된 바와 같이, 동적-정적 복셀 특징 X의 각 정적 복셀 특징 x_i를 LSP 특징 E로 인코딩한다. 여기서 동적-정적 복셀 특징 X에 포함된 각 정적 복셀 특징 x_i에 포함된 3차원 픽셀의 개수는 정규화되어 있을 수 있다. 동적-정적 복셀 특징 X의 각 정적 복셀 특징 x_i를 LSP 특징 E로 인코딩하는 방법에 대해서는 도 4 및 도 5를 참조하여 설명한다.

[0067] 도 4 및 도 5는 정적 복셀 특징을 LSP 특징으로 인코딩하는 방법을 설명하기 위한 도면이다.

[0068] LSP 특징 E는 사각 큐브를 구성하는 복수의 복셀 플레인(voxel plane) 각각의 내부에 3차원 픽셀들이 포함된 정도를 나타낸다. LSP 특징 E은 복수의 복셀 플레인 각각의 내부에 3차원 픽셀들이 포함된 정도를 나타내는

LSP 엘리먼트들을 포함할 수 있다.

[0069] 전술한 바와 같이, 3D 손바닥 자세 추정기 $\mathscr{R}$ 에 의해 손바닥의 회전 R 및 이동 T를 나타내는 변환 정보 M가 획득되면, 복수의 프레임 각각에서 손의 손바닥의 지향(orientation) 방향이 결정되고, 이에 따라 소정 지점을 원점으로 하였을 때의 로컬 좌표의 각 축(x축, y축 및 z축)의 방향이 결정될 수 있다. 사각 큐브를 구성하는 복수의 복셀 플레인에 대해서는 로컬 좌표 상에서의 인덱스가 매핑되고, 어느 하나의 위치에 매핑된 복셀 플레인에 대응하는 LSP 엘리먼트들이 추출되어 LSP 특징에 포함될 수 있다. 사각 큐브를 구성하는 복셀 플레인은 $\mathrm{LSP}^{\hat{u},j}$ 로 표현될 수 있는데, 여기서, $\hat{u}$ 는 로컬 좌표 상의 특정 축을 나타내고, j 는 특정 축에서의 인덱스를 나타낸다. 또한, 복셀 플레인 $\mathrm{LSP}^{\hat{u},j}$ 에 대응하는 LSP 엘리먼트는 $(\mathrm{LSP}^{\hat{u},j})_{axis}$ 로 표현될 수 있으며, 여기서 axis는 복셀 플레인 $\mathrm{LSP}^{\hat{u},j}$ 의 LSP 값을 추출하기 위한 계산 방향을 나타낸다. 예를 들어, axis가 y인 경우, $\mathrm{LSP}^{\hat{u},j}$ 의 LSP 값을 y축의 방향으로 계산하였다는 것을 의미한다.

[0070] 도 5(a)는 x축의 인덱스 2의 위치에 매핑된 복셀 플레인(500)을 도시하고 있다. 도 5(b)는 어느 하나의 프레임에서 x축의 인덱스 2의 위치에 매핑된 복셀 플레인(500)을 도시하고 있고, 도 5(c)는 다른 하나의 프레임에서 x축의 인덱스 2의 위치에 매핑된 복셀 플레인(500)을 도시하고 있다. 도 5(b) 및 도 5(c)에서 음영 부분은 3차원 픽셀이 위치하는 영역을 가리킨다.

[0071] 제스쳐 감지 장치는 x축의 인덱스 2의 위치에 매핑된 복셀 플레인(500)의 대응하는 y축 LSP 엘리먼트와 z축 LSP 엘리먼트를 추출한다. 구체적으로, 제스쳐 감지 장치는 x축의 인덱스 2의 위치에 매핑된 복셀 플레인(500)이 y축을 따라 n개의 층으로 구분되어 있는 경우, 각 층의 복셀 내의 3차원 픽셀의 밀도에 대응하여 상기 n개의 층 각각에 할당된 값이 나열된 값을 y축 LSP 엘리먼트로 추출하고, x축의 인덱스 2의 위치에 매핑된 복셀 플레인(500)이 z축을 따라 m개의 층으로 구분되어 있는 경우, 각 층의 복셀 내의 3차원 픽셀의 밀도에 대응하여 상기 m개의 층 각각에 할당된 값이 나열된 값을 z축 LSP 엘리먼트로 추출한다. 도면으로 도시되지는 않았으나, y축 상에 위치하는 복셀 플레인에 대해서는 x축 LSP 엘리먼트와 z축 LSP 엘리먼트가 추출되고, z축 상에 위치하는 복셀 플레인에 대해서는 x축 LSP 엘리먼트와 y축 LSP 엘리먼트가 추출된다. 사각 큐브를 구성하는 모든 복셀 플레인에 대해 LSP 엘리먼트들이 추출될 수 있다.

[0072] 도 5(b)에 도시된 바와 같이, x축의 인덱스 2의 위치에 매핑된 복셀 플레인(500)이 y축을 따라 5개의 층으로 구성되어 있는 경우, 각 층에서의 3차원 픽셀의 밀도를 계산하여 그에 대응하는 값을 할당한다. 도 5(b)에서는 1층부터 5층까지 모두 1이 할당된 것을 알 수 있다. 제스쳐 감지 장치는 원점에서 가장 멀리 떨어진 층에 할당된 값이 가장 높은 자릿수를 가지며, 원점에 가까워질수록 자릿수가 작아지도록 각 층에 할당된 값을 나열한다. 이에 따라, 제스쳐 감지 장치는 y축 LSP 엘리먼트를 11111로 결정한다.

[0073] 또한, 도 5(b)에서 x축의 인덱스 2의 위치에 매핑된 복셀 플레인(500)이 z축을 따라 3개의 층으로 구성되어 있는 경우, 각 층에서의 3차원 픽셀의 밀도를 계산하여 그에 대응하는 값을 할당한다. 도 5(b)에서는 1층에 5, 2층에 0, 3층에 0이 할당된 것을 알 수 있다. 제스쳐 감지 장치는 원점에서 가장 멀리 떨어진 층에 할당된 값이 가장 높은 자릿수를 가지며, 원점에 가까워질수록 자릿수가 작아지도록 각 층에 할당된 값을 나열한다. 이에 따라, 제스쳐 감지 장치는 z축 LSP 엘리먼트를 005로 결정한다.

[0074] 제스쳐 감지 장치는 3차원 픽셀에 밀도에 대응하는 값을 각 층에 할당할 때, 8진수 시스템(octal number system)에 의할 수 있다. 즉, 3차원 픽셀의 밀도가 클수록 큰 수를 각 층에 할당하되, 최대 값은 7로 제한하는 것이다. 이에 의하면, 도 5(b)의 y축 LSP 엘리먼트 11111은 10진수로 변환되어 4681의 값을 갖고, z축 LSP 엘리먼트 005는 10진수로 변환되어 5의 값을 갖게 된다.

[0075] 다음으로, 도 5(c)에서 x축의 인덱스 2의 위치에 매핑된 복셀 플레인(500)이 y축을 따라 5개의 층으로 구성되어 있는 경우, 각 층에서의 3차원 픽셀의 밀도를 계산하여 그에 대응하는 값을 할당한다. 도 5(c)에서는 1층에 1, 2층에 2, 3층에 2, 4층 및 5층에 0이 할당된 것을 알 수 있다. 이에 따라, 제스쳐 감지 장치는 y축 LSP 엘리먼트를 00221로 결정할 수 있다. 또한, 복셀 플레인(500)이 z축을 따라 3개의 층으로 구성되어 있는 경우, 도 5(c)와 같이 1층에 3, 2층에 1 및 3층에 1을 할당할 수 있으며, 이에 따라, 제스쳐 감지 장치는 z축 LSP 엘리먼

트를 113으로 결정한다. 제스쳐 감지 장치는 y축 LSP 엘리먼트로 00221을 결정하고, z축 LSP 엘리먼트로 113을 결정할 수도 있으나, 이들을 10진수로 변환한 값인 145와 75 각각을 y축 LSP 엘리먼트와 z축 LSP 엘리먼트로 결정할 수도 있다.

[0076] 도 4를 보면, 제스쳐 감지 장치는 복수의 프레임 각각에 대응하는 정적 복셀 특징 x_i를 인코딩하여 복수의 LSP 특징($E_{[0:m]}$)을 획득할 수 있다. 하나의 사각 큐브 내에 10 * 10 * 10의 복셀이 존재하는 경우, 정적 복셀 특징 x_i의 엘리먼트는 1000개에 해당하나, LSP 특징의 LSP 엘리먼트는 60개 (10*2 + 10*2 + 10*2)로 축소될 수 있다.

[0077] 제스쳐 감지 장치는 하나의 프레임에서 추출된 복수의 LSP 엘리먼트들을 기 설정된 순서로 나열하고, 나열된 LSP 엘리먼트들에 인덱스를 부여할 수 있다. LSP 엘리먼트의 변수를 $\hat{u}$, j, axis라 할 때, 예를 들어, LSP 엘리먼트들에 대한 인덱스는 다음과 같이 부여될 수 있다.

[0078] [Index: $\hat{u}$, j, axis], [0: x, 0, Y], [1: x, 0, Z], [2: x, 1, Y], [3: x, 1, Z], [4: y, 0, X], [5: y, 0, Z], [6: y, 1, X], [7: y, 1, Z], [8: z, 0, X], [9: z, 0, Y], [10: z, 1, X], [11: z, 1, Y]

[0079] 이상과 같이, 본 발명의 일 실시예에 따른 제스쳐 감지 장치는 복수의 프레임을 포함하는 제 1 시퀀스 영상이 획득되면, 제 1 시퀀스 영상으로부터 복수의 제 1 LSP 특징을 추출하고, 추출된 제 1 LSP 특징에 기초하여, 제 1 시퀀스 영상의 손의 홀딩 제스쳐에 대응하는 물체의 종류와 조작 제스쳐에 대응하는 조작의 종류를 판단할 수 있다. 제스쳐 감지 장치는 홀딩 제스쳐에 대응하는 물체의 종류와 조작 제스쳐에 대응하는 조작의 종류를 판단하기 위해 랜덤 포레스트를 미리 구성해 놓을 수 있는데, 이에 대해서는 도 6을 참조하여 설명한다.

[0080] 도 6은 손의 제 2 시퀀스 영상의 복수의 프레임 각각으로부터 추출되어 인코딩된 제 2 LSP 특징에 기초하여 구성된 랜덤 포레스트의 트리(600)를 나타내는 예시적인 도면이다. 랜덤 포레스트는 제스쳐 감지 장치에 의해 구성될 수도 있지만, 미리 구성되어 제스쳐 감지 장치에 저장될 수도 있다. 이하에서는, 제스쳐 감지 장치가 랜덤 포레스트를 구성할 때의 방법을 설명한다.

[0081] 먼저, 제스쳐 감지 장치는 홀딩 제스쳐에 대응하는 물체의 종류(a)와 조작 제스쳐에 대응하는 조작의 종류(a')를 미리 알고 있는 손의 제 2 시퀀스 영상으로부터 복수의 제 2 LSP 특징을 획득할 수 있다. 홀딩 제스쳐에 대응하는 물체는 예를 들어, 볼펜, 스마트폰, 스프레이 등의 다양한 물체를 포함할 수 있으며, 조작 제스쳐에 대응하는 조작은 예를 들어, 프레싱(pressing), 중립(neutral), 릴리즈(release) 등을 포함할 수 있으나, 이에 한정되지는 않는다.

[0082] 랜덤 포레스트는 이진 결정 트리(binary decision tree)의 앙상블(ensemble)로서, 각각의 트리는 두 가지 종류의 노드, 즉, 스플릿 노드(split node)와 리프 노드(leaf node)를 포함한다. 본 발명에서의 랜덤 포레스트의 각 스플릿 노드들은 입력 데이터에 대하여, 특정 업무의 분할 함수를 수행하고, 이를 왼쪽의 하위 노드로 라우팅할지, 오른쪽의 하위 노드로 라우팅할지를 결정한다. 리프 노드는 물체의 종류 및 조작의 종류를 나타내는 종단 노드로서 물체의 종류의 확률과 조작의 종류의 확률을 저장한다.

[0083] 랜덤 포레스트 내의 각 트리는 현재의 훈련(또는 학습) 데이터를 재귀적으로(recursively) 분할하고, 두 개의 자손 노드(child node)로 보내면서 성장한다. 트리의 각 노드에서는 무작위로 분할 후보들 $\Psi = (f_\iota, \tau_\iota)$이 생성되며, 여기서, f_ι은 분할 함수이고, τ_ι은 임계 값으로서, 제스쳐 감지 장치는 스플릿 노드에서 가장 큰 정보 게인(information gain: IG)을 제공하는 최적의 파라미터 세트인 $\psi^\dagger_\iota \in \Psi$를 찾는다.

[0084] 본 발명의 일 실시예에서, 각 스플릿 노드에서의 분할 함수 f_ι은 아래의 수학식 4에 따라 결정된다.

[0086] [수학식 4]

$$f_\iota(\mathrm{E}) = \begin{cases} (E_{m-p})_{idx}, \text{ if layer:top} \\ (E_{m-p})_{idx} - (E_{m-q})_{idx}, \text{ if layer:bottom} \end{cases}$$

[0087]

[0088] 수학식 4에서 $f_\iota(\mathrm{E})$ 는 분할 함수, $(E_{m-p})_{idx}$는 정적 함수로서 LSP 특징에 포함된 LSP 엘리먼트, $(E_{m-p})_{idx} - (E_{m-q})_{idx}$는 동적 함수로서, $(E_{m-q})_{idx}$는 LSP 특징에 포함된 LSP 엘리먼트, p 및 q는 시퀀스 영상의 마지막 m 프레임으로부터의 선행 오프셋, idx는 LSP 특징에 포함된 LSP 엘리먼트들 중 어느 하나를 나타내는 인덱스를 나타낸다.

[0090] 본 발명의 일 실시예에서 제스처 감지 장치는 트리에서 상위 계층(top layer)에 위치하는 스플릿 노드에 대해서는 정적 함수를 할당하고, 하위 계층(bottom layer)에 위치하는 스플릿 노드에 대해서는 동적 함수를 할당한다. 즉, 상위 계층에서 정적 함수에 의해 홀딩 제스처에 대응하는 물체의 종류가 어느 정도 확정된 뒤에, 하위 계층에서 조작 제스처에 대응하는 조작의 종류를 판단하는 것이다.

[0091] 구체적으로, 제스처 감지 장치는 트리 내 최상위 층의 스플릿 노드의 분할 함수를 정적 함수로 결정하고, 정적 함수의 파라미터(p, idx)를 무작위로 선택하면서 입력되는 데이터 세트를 분할하여 하위 노드로 전달한다. IG가 최대가 되는 파라미터가 선택되면, 제스처 감지 장치는 최상위 계층의 분할 함수의 파라미터 최적화를 완료하고, 학습 데이터를 하위 노드로 전달한다. 최상위 층에 위치하는 스플릿 노드의 분할 함수는 항상 정적 함수로 결정된다.

[0092] 분할 함수 f_ι 과 무작위로 선택된 임계 값 τ_ι 에 기초하여, 주어진 데이터 세트 $\mathscr{D}$가 서브 세트인 $\mathscr{D}_l$와 $\mathscr{D}_r$ 로 분할되되, $\mathscr{D}_l = \{(E_{m-p})_{idx} < \tau_\iota\}$ 이고, $\mathscr{D}_r = \mathscr{D} \setminus \mathscr{D}_l$ 이다. 그리고, IG는 아래의 수학식 5에 따라 결정된다.

[0094] [수학식 5]

$$IG(\mathscr{D}_o, \psi^\dagger{}_\iota) = H(\mathscr{D}) - \sum_{e\in\{l,r\}} \frac{|\mathscr{D}^e|}{|\mathscr{D}|} H(\mathscr{D}^e)$$

[0095]

[0097] 상기 수학식 5에서 H(·)은 새넌 엔트로피(shannon's entropy)로서, 아래의 수학식 6에 따라 결정된다.

[0099] [수학식 6]

$$H(\mathscr{D}) = -\sum_{\forall g\in\{G^\dagger\}} p(g|\mathscr{D}) log(p(g|\mathscr{D}))$$

[0100]

[0102] 상기 수학식 5 및 수학식 6에서 $p(g|\mathscr{D})$는 데이터 세트 $\mathscr{D}$ 내에서 g로 라벨링된 제스처의 확률, $G^\dagger$는 제스처 타입(예를 들어, 홀딩 제스처 및 조작 제스처)과 관련된 현재 층(layer)에 의존하는 제스처 라벨의 집합을 나타낸다. 예를 들어, $G^\dagger$ 가 상위 계층에서 복수의 홀딩 제스처 라벨 {0:N}의 집합을 나타낸다면, 라벨 g는 0, 1, 2, …, N 중에서의 하나의 라벨에 해당할 것이고, $G^\dagger$ 가 하위 계층에서의 복수의 조작 제스처 라벨 {P, N, R}의 집합을 나타낸다면, 라벨 g는 P(pressing), N(neutral) 및 R(release) 중에서의 하나의 라벨에 해당할 것이다.

[0103] 최상위 계층의 하위에 위치하는 스플릿 노드에서는, 전달되는 데이터 세트에 포함된 데이터들의 대부분이 어느

하나의 물체의 종류에 대응하는 홀딩 제스처를 나타내는 것으로 라벨링(labelling)되어 있는지에 따라, 분할 함수가 정적 함수 또는 동적 함수로 결정된다. 다시 말하면, 최상의 층의 스플릿 노드에서 나뉘어진 데이터 세트를 입력받는 하위층의 스플릿 노드에서는, 데이터 세트 중 동일한 제스처 라벨로 라벨링된 데이터의 데이터 세트 내에서의 가장 큰 확률 (또는 비율)과 두 번째로 큰 확률 (또는 비율) 사이의 차이가 임계값보다 큰 경우, 분할 함수를 동적 함수로 결정하고, 그렇지 않다면, 분할 함수를 정적 함수로 결정하는 것이다. 상기 임계값은 예를 들어, 0.9로 결정될 수 있다. 예를 들어, 데이터 세트 중 스프레이로 라벨링된 데이터의 확률이 가장 크고, 스마트폰으로 라벨링된 데이터의 확률이 두 번째로 크고, 스프레이로 라벨링된 데이터의 확률과 스마트폰으로 라벨링된 데이터의 확률 사이의 차이가 임계값보다 큰 경우, 해당 데이터 세트의 대부분의 데이터는 스프레이를 홀딩하고 있는 것이므로, 제스처 감지 장치는 해당 스플릿 노드에서 더 이상 홀딩 제스처를 판단할 필요가 없다고 결정하고, 분할 함수를 동적 함수로 결정하는 것이다.

[0104] 데이터 세트 중 동일한 물체를 홀딩하고 있는 것으로 라벨링된 데이터가 많다는 것은 상위 계층의 스플릿 노드에서 물체의 종류의 판단이 거의 완료되었다는 것을 의미하기 때문에, 하위 계층에서는 더 이상 물체의 종류를 판단하지 않는다.

[0105] 도 6을 보면, 최상위 계층의 스플릿 노드(610)에서의 분할 함수는 정적 함수로 결정되고, 두 번째 계층의 좌측 스플릿 노드(620)에서의 분할 함수는 동적 함수로, 우측 스플릿 노드(630)에서의 분할 함수는 정적 함수로 결정되는 것을 알 수 있다. 또한, 두 번째 계층의 좌측 스플릿 노드(620)에서의 분할 함수가 동적 함수로 결정된 이후에는, 하위 계층의 모든 스플릿 노드의 분할 함수가 동적 함수로 결정된다는 것을 알 수 있다.

[0106] 제스처 감지 장치는 동적 함수로 결정된 스플릿 노드에 대해서도 파라미터(p, q, idx)를 무작위로 선택하면서 입력되는 데이터 세트를 분할하여 하위 노드로 전달한다. IG가 최대가 되는 파라미터가 선택되면, 제스처 감지 장치는 분할 함수의 파라미터 최적화를 완료하고, 학습 데이터를 하위 노드로 전달한다. 동적 함수에서는 서로 다른 프레임에 대응하는 서로 다른 LSP 특징의 LSP 엘리먼트를 비교하기 때문에, 손의 특정 부분의 이동 경향이 판단될 수 있다.

[0107] 트리의 성장은 중단 기준에 만족할 때까지 재귀적으로 반복된다. 중단 기준은 예를 들어, 데이터 세트의 샘플이 개수가 기 설정된 최소 개수보다 작거나, 트리의 깊이(depth)가 기 설정된 값을 초과하는 경우나, 현재의 노드로 전달된 분할 데이터 세트의 모든 조작 제스처의 라벨들이 동일한 라벨에 해당하는 경우일 수 있다.

[0108] 트리의 각 리프 노드에는 홀딩 제스처의 확률과 조작 제스처의 확률이 저장된다.

[0109] 도 5를 참조하여 설명하였던 제 1 LSP 특징들이 이진 트리에 입력되면, 물체의 종류에 대응하는 홀딩 제스처의 확률과, 조작의 종류에 대응하는 조작 제스처의 확률이 아래의 수학식 7에 따라 결정된다.

[0111] [수학식 7]

$$\alpha^*, \beta^* = \underset{\alpha \in \mathrm{A}, \beta \in \mathrm{B}}{\arg\max} P(Y_\alpha, Z_\beta | \mathrm{E})$$

[0112]

[0113] 상기 수학식 7에서, $P(Y_\alpha)$ 는 홀딩 제스처 α 의 확률, A는 홀딩 제스처들의 집합, $P(Z_\beta)$ 는 조작 제스처 β 의 확률, B는 조작 제스처들의 집합을 나타낸다.

[0114] 홀딩 제스처 Y_α 의 확률과, 조작 제스처 Z_β 의 확률은 랜덤 포레스트 내 트리들의 리프 노드에 저장된 값을 평균함으로써, 아래의 수학식 8과 같이 결정된다.

[0116] [수학식 8]

$$P(Y_\alpha, Z_\beta | \mathrm{E}) = \frac{1}{T} \sum_{\forall tree \in T} P_{tree}(Y_\alpha, Z_\beta | \mathrm{E})$$

[0117]

[0118] 제스처 감지 장치는 랜덤 포레스트의 복수의 트리 각각에서 결정된 물체의 종류와 조작의 종류의 확률의 평균과 기 설정된 임계 값을 비교하여, 홀딩 제스처에 대응하는 물체의 종류 및 조작 제스처에 대응하는 조작의 종류를

최종적으로 판단할 수 있다.

[0119] 이하에서는, 도 7 및 도 8을 참조하여 본 발명의 일 실시예에 따른 제스쳐 감지 방법의 효과를 살펴본다.

[0120] 도 7(a) 및 도 7(b) 각각은 다양한 시점(view point)에서 관측된 데이터 세트를 참고 문헌 [6] 및 [7]의 종래 기술에 적용한 경우와, 본 발명의 일 실시예에 따른 제스쳐 감지 방법에 적용한 경우, 물체의 종류의 판단 정확도를 비교하기 위한 그래프이다.

[0121] 도 7(a)을 참조하면, 종래 기술에 의한 경우, 물체 종류의 판단 정확도가 60.00%를 보이는 반면에, 도 7(b)를 참조하면, 본 발명의 제스쳐 감지 방법에 의한 경우, 정확도가 91.36%까지 향상된 것을 알 수 있다.

[0122] 다음으로, 도 8은 데이터 세트가 입력되었을 때, 참고 문헌 [8]에 따른 조작의 종류의 판단 정확도와, 제스쳐 감지 방법에 따른 조작의 종류의 판단 정확도를 비교하기 위한 ROC(Receiver Operating Characteristic) 그래프이다. 도 8은 엄지 손가락에 대한 조작 종류의 판단 정확도를 나타내는데, 종래 기술에 따르면 89.80%의 정확도를 보이는데 반해, 본 발명에 따르면, 95.59%의 정확도를 보이고 있다.

[0123] 도 9는 본 발명의 다른 실시예에 따른 제스쳐 감지 장치(900)의 구성을 나타내는 블록도이다.

[0124] 도 9를 참조하면, 본 발명의 다른 실시예에 따른 제스쳐 감지 장치(900)는 영상 획득부(910), 영상 처리부(930), 특징 추출부(950) 및 판단부(970)를 포함할 수 있다. 영상 획득부(910), 영상 처리부(930), 특징 추출부(950) 및 판단부(970)는 적어도 하나의 마이크로 프로세서로 구현될 수 있으며, 도시되지 않은 메모리에 저장된 프로그램에 따라 동작할 수 있다.

[0125] 본 발명의 다른 실시예에 따른 제스쳐 감지 장치(900)는 사용자에게 착용 가능한 장치로서, 깊이 카메라를 더 포함할 수도 있다. 또는, 구현예에 따라서는, 제스쳐 감지 장치(900)는 사용자에게 착용된 깊이 카메라와는 별개의 컴퓨터로서 구현될 수도 있다.

[0126] 영상 획득부(910)는 사용자에게 착용된 깊이 카메라에 의해 촬영되는 손의 제 1 시퀀스 영상을 획득한다. 또한, 영상 획득부(910)는 랜덤 포레스트 구성을 위한 제 2 시퀀스 영상을 획득할 수 있다.

[0127] 영상 처리부(930)는 제 1 시퀀스 영상 내 복수의 프레임 각각의 손에 복수의 복셀로 나뉘어진 사각 큐브를 매핑한다.

[0128] 특징 추출부(950)는 제 1 시퀀스 영상 내 복수의 프레임 각각에 대해, 사각 큐브를 구성하는 복수의 복셀 각각의 내부에 포함된 손의 3차원 픽셀들의 개수를 나타내는 정적 복셀 특징을 추출하고, 추출된 정적 복셀 특징을 인코딩하여 제 1 LSP 특징을 획득한다.

[0129] 판단부(970)는 제 1 LSP 특징을 랜덤 포레스트에 입력하여, 제 1 시퀀스 영상에서 손의 홀딩 제스쳐에 대응하는 물체의 종류와 조작 제스쳐에 대응하는 조작의 종류를 판단할 수 있다. 상기 랜덤 포레스트는 홀딩 제스쳐에 대응하는 물체의 종류 및 조작 제스쳐에 대응하는 조작의 종류를 알고 있는 손의 제 2 시퀀스 영상의 복수의 프레임 각각으로부터 획득된 제 2 LSP 특징에 기초하여 구성될 수 있다.

[0130] 한편, 상술한 본 발명의 실시예들은 컴퓨터에서 실행될 수 있는 프로그램으로 작성가능하고, 작성된 프로그램은 매체에 저장될 수 있다.

[0131] 상기 매체는 마그네틱 저장매체(예를 들면, 롬, 플로피 디스크, 하드디스크 등), 광학적 판독 매체(예를 들면, 시디롬, 디브이디 등) 및 캐리어 웨이브(예를 들면, 인터넷을 통한 전송)와 같은 저장매체를 포함할 수 있으나, 이에 한정되는 것은 아니다.

[0132] 첨부된 도면을 참조하여 본 발명의 실시예를 설명하였지만, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자는 본 발명이 그 기술적 사상이나 필수적인 특징을 변경하지 않고서 다른 구체적인 형태로 실시될 수 있다는 것을 이해할 수 있을 것이다. 그러므로 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적이 아닌 것으로 이해해야만 한다.

부호의 설명

[0133] 900: 제스쳐 감지 장치

910: 영상 획득부

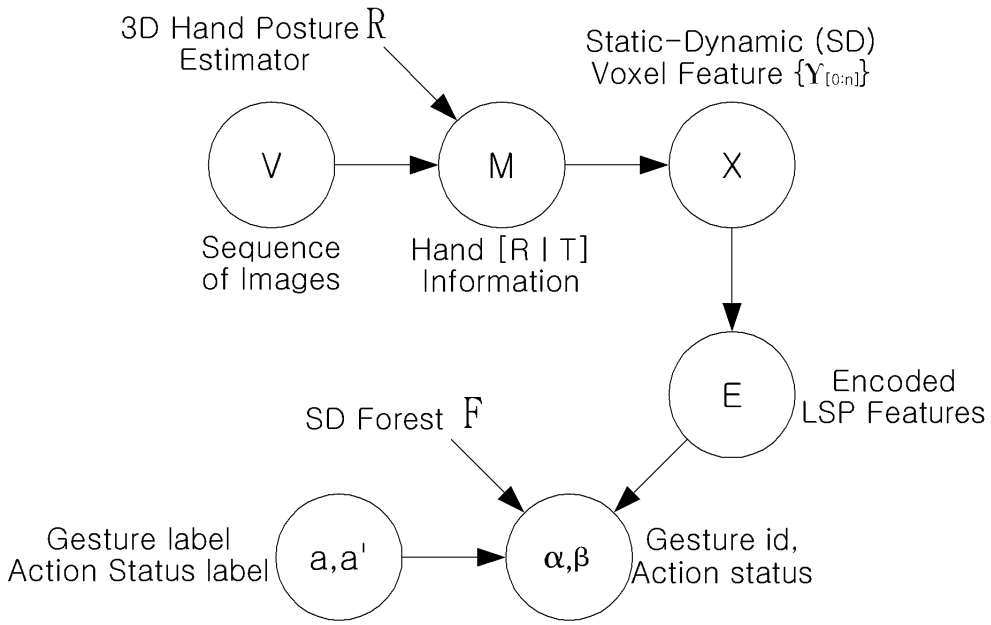
930: 영상 처리부

950: 특징 추출부

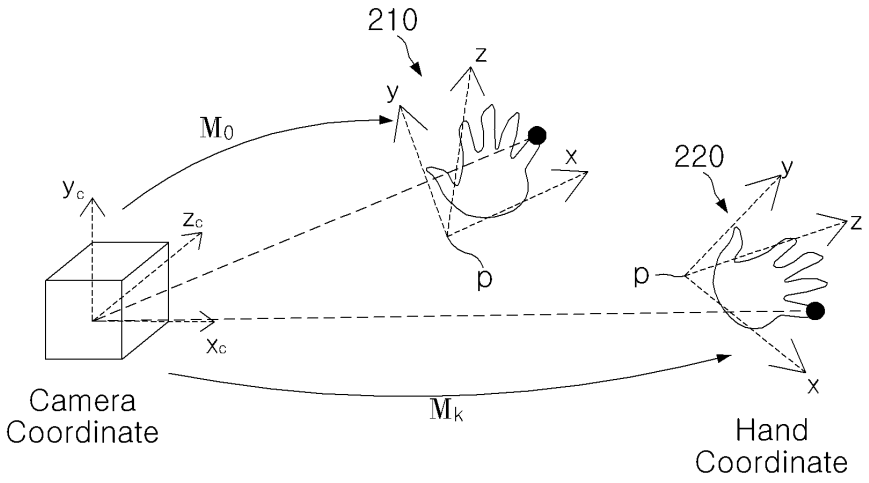
970: 판단부

도면

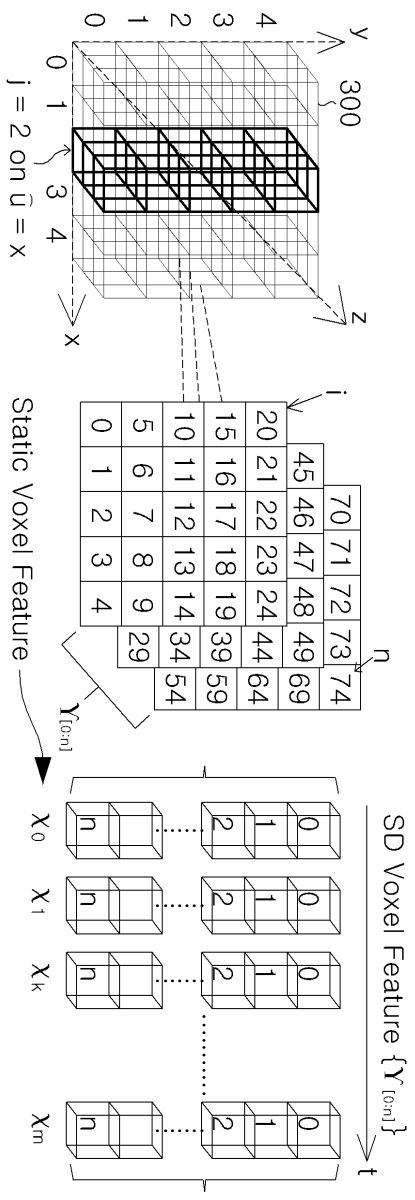
도면1



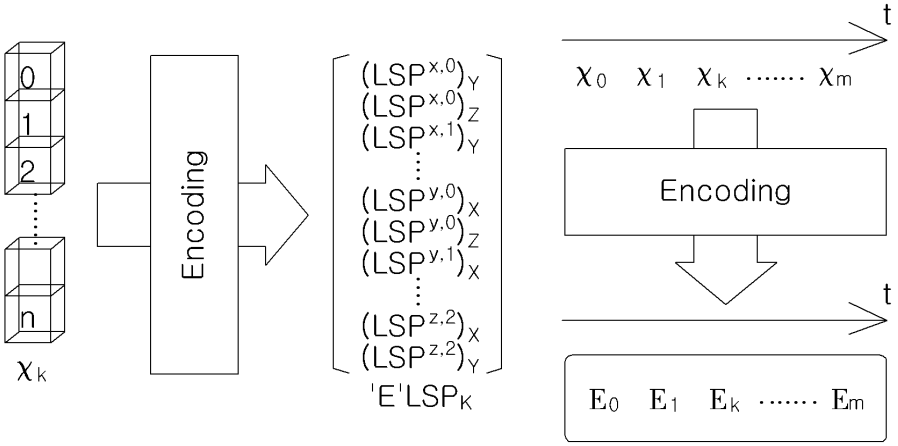
도면2



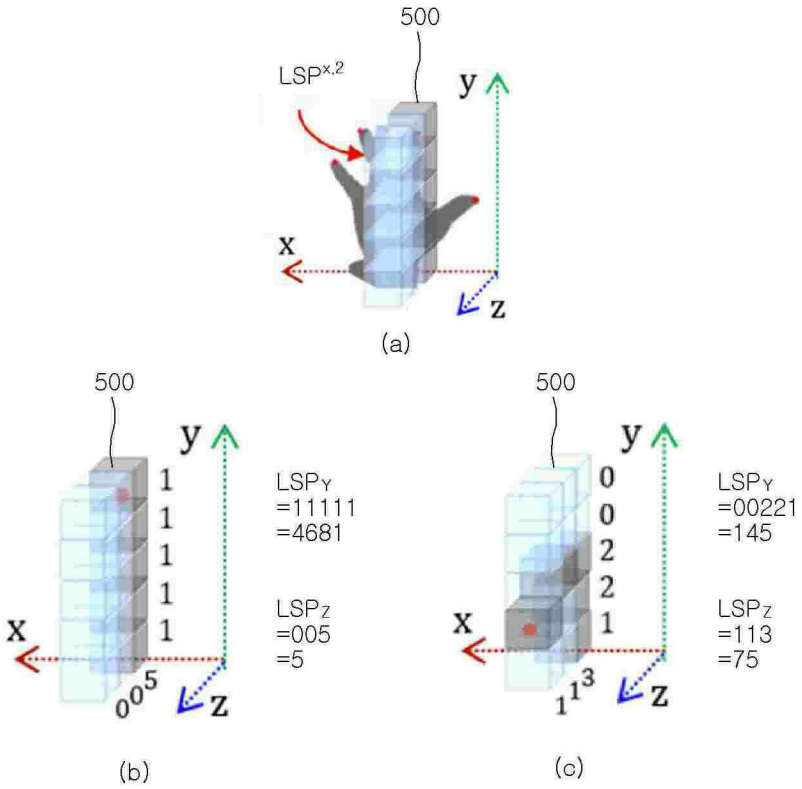
도면3


Static Voxel Feature
SD Voxel Feature {Y[0:n]}
j = 2 on u = x
300
i
n
t

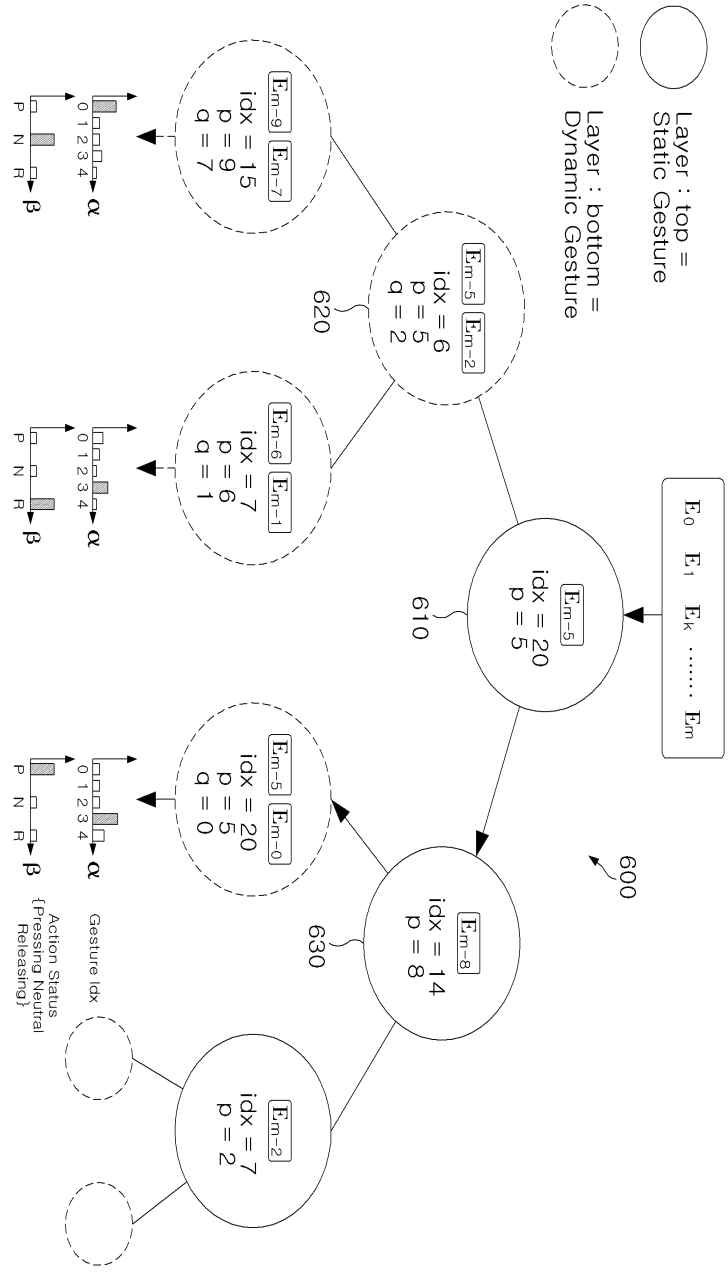
도면4


Encoding
'E'LSPK
Encoding
t
t

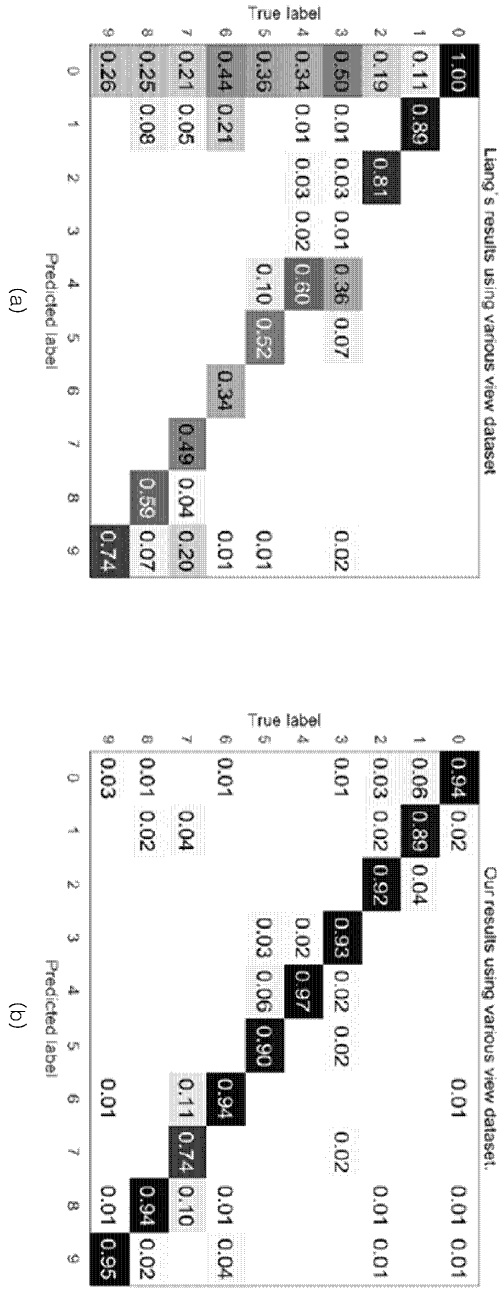
도면5


500
LSP^x,2
y
x
z
(a)
500
y
1
1
1
1
1
LSPY
=11111
=4681
LSPZ
=005
=5
x
0 0 5
z
(b)
500
y
0
0
2
2
1
LSPY
=00221
=145
LSPZ
=113
=75
x
1 1 3
z
(c)

도면6


Layer : top = Static Gesture
Layer : bottom = Dynamic Gesture
E0 E1 Ek ..... Em
600
Em-5
idx = 20
p = 5
610
Em-8
idx = 14
p = 8
630
Em-5 Em-2
idx = 6
p = 5
q = 2
620
Em-9 Em-7
idx = 15
p = 9
q = 7
Em-6 Em-1
idx = 7
p = 6
q = 1
Em-5 Em-0
idx = 20
p = 5
q = 0
Em-2
idx = 7
p = 2
P N R
0 1 2 3 4
α
β
Gesture Idx
Action Status
{Pressing Neutral
Releasing}

도면7

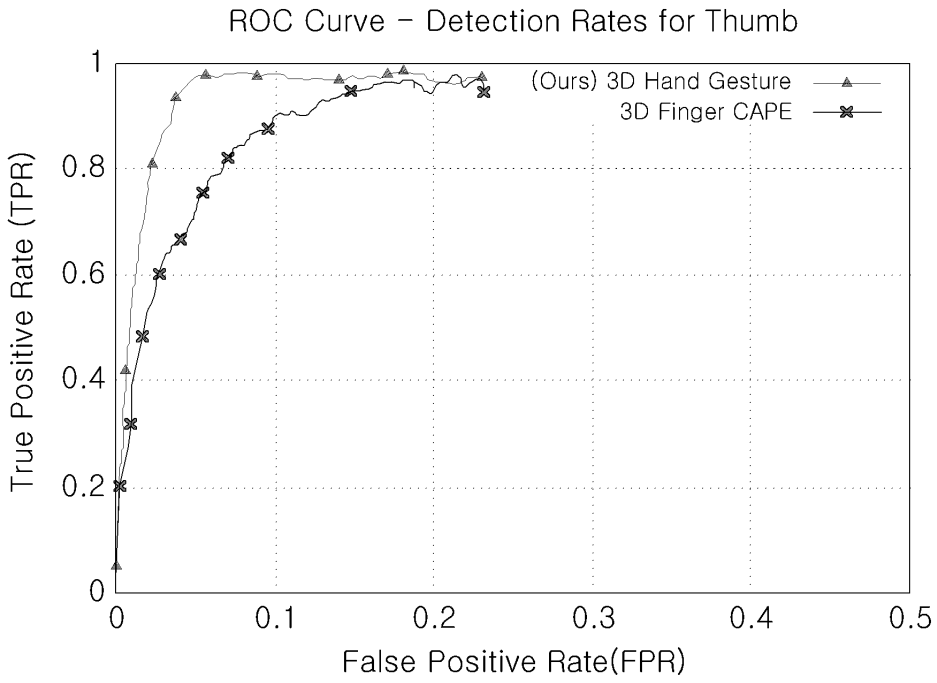

Liang's results using various view dataset
True label
Predicted label
0 1.00
1 0.11 0.89
2 0.19 0.81
3 0.50 0.01 0.03 0.01 0.36 0.07 0.02
4 0.34 0.01 0.03 0.02 0.60
5 0.36 0.10 0.52 0.01
6 0.44 0.21 0.34 0.01
7 0.21 0.05 0.49 0.04 0.20
8 0.25 0.08 0.59 0.07
9 0.26 0.74
0 1 2 3 4 5 6 7 8 9

(a)

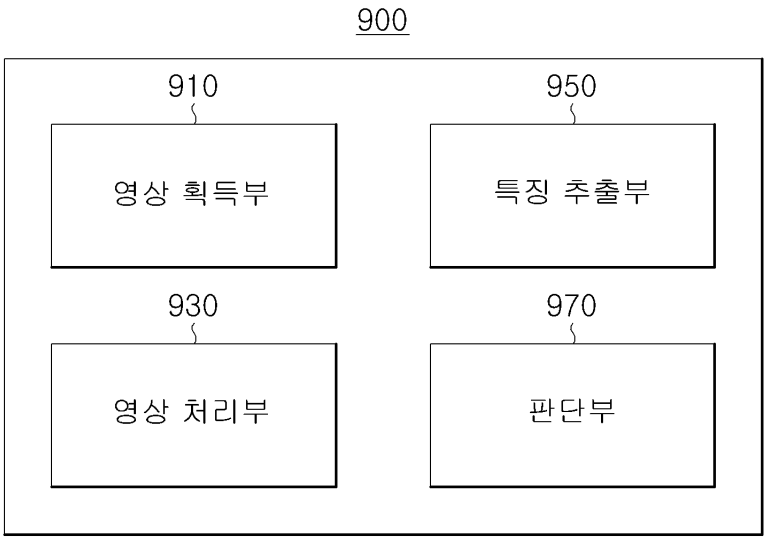
Our results using various view dataset.
True label
Predicted label
0 0.94 0.02 0.01 0.01 0.01
1 0.06 0.89 0.04
2 0.03 0.02 0.92 0.01 0.01
3 0.01 0.93 0.02 0.02
4 0.02 0.97 0.03 0.06 0.90
5 0.02 0.94
6 0.01 0.11 0.74 0.10 0.01 0.04
7 0.04
8 0.01 0.02 0.94 0.02
9 0.03 0.01 0.01 0.95
0 1 2 3 4 5 6 7 8 9

(b)

도면8


ROC Curve - Detection Rates for Thumb
(Ours) 3D Hand Gesture
3D Finger CAPE
True Positive Rate (TPR)
False Positive Rate(FPR)
0
0.2
0.4
0.6
0.8
1
0.1
0.3
0.5

도면9


900
910
영상 획득부
950
특징 추출부
930
영상 처리부
970
판단부