특집논문 (Special Paper)
방송공학회논문지 제17권 제3호, 2012년 5월 (JBE Vol. 17, No. 3, May 2012)
http://dx.doi.org/10.5909/JBE.2012.17.3.447

키넥트 센서 데이터를 이용한 손 제스처 인식

조 선 영[a], 변 혜 란[a]‡, 이 희 경[b], 차 지 훈[b]

Hand Gesture Recognition from Kinect Sensor Data

Sunyoung Cho[a], Hyeran Byun[a]‡, Hee Kyung Lee[b], and Jihun Cha[b]

요 약

본 논문에서는 키넥트 센서로부터 획득한 관절 정보를 이용하여 손 제스처를 인식하는 방법을 나타낸다. 관절 정보에 대한 관찰열을 표현하기 위한 특징으로 방향 변형에 강인한 다각도 결합 히스토그램 특징을 제안한다. 제안한 특징은 다양한 각도의 양자화 레벨을 갖는 여러 개의 각도 히스토그램들을 결합함으로써, 사람 및 환경에 따라 발생할 수 있는 제스처의 방향 변형에 강인하게 제스처를 표현한다. 또한, 다각도 결합 히스토그램으로 표현된 제스처 관찰열은 랜덤 결정 포레스트 분류기와 잘 결합되어 높은 성능으로 제스처의 클래스를 인식한다. 키넥트 센서로부터 획득한 정적 및 동적 타입의 손 제스처 데이터셋에서 실험을 진행하였고, 다른 제스처 특징 및 분류기를 갖는 방법과의 인식 성능 비교를 통해 제안하는 방법의 우수함을 입증하였다.

Abstract

We present a method to recognize hand gestures using skeletal joint data obtained from Microsoft's Kinect sensor. We propose a combination feature of multi-angle histograms robust to orientation variations to represent the observation sequence of skeletons. The proposed feature efficiently represents the orientation variations of gestures that can be occurred according to person or environment by combining the multiple angle histograms with various angular-quantization levels. The gesture represented as combination of multi-angle histograms and random decision forest classifier improve the recognition performance. We conduct the experiments in hand gesture dataset obtained from a kinect sensor and show that our method outperforms the other methods by comparing the recognition performance.

Keyword : Hand gesture recognition, Combination of multi-angle histograms, Kinect

Ⅰ. 서 론

손 제스처 인식은 많은 인간-컴퓨터 상호작용(Human-Computer Interaction, HCI) 응용에서 요구되는 기술로써 현재까지 활발히 연구되고 있다. 손 제스처를 인식하기 위해서는 크게 2가지 문제를 해결해야 한다: 1) 손 제스처를

a) 연세대학교 컴퓨터과학과 (Dept. of Computer Science, Yonsei University)
b) 한국전자통신연구원 방통융합미디어연구부(Broadcasting & Telecommunications Convergence Media Research Dept., ETRI)
‡ 교신저자 : 변혜란 (Hyeran Byun)
E-mail: hrbyun@yonsei.ac.kr
Tel: +82-2-2123-2719, Fax: +82-2-363-2599
· 접수일(2012년3월15일),수정일(2012년4월20일),게재확정일(2012년4월23일)

잘 표현하는 특징을 선택하여 추출하고, 2) 추출된 특징을 이용하여 제스처 클래스로 분류한다. Murthy와 Jadon[1]은 손 제스처를 인식하기 위한 특징들을 3가지 접근방법으로 분류하였다. 첫 번째는 손의 포즈 정보를 추정하는 모델 기반 방법이고, 두 번째는 영상 시퀀스로부터 손 제스처를 모델링하는 뷰(View) 기반 방법이다. 세 번째는 손 영역의 위치나 움직임과 같은 특징을 사용하는 저차원 특징 기반 방법이다. 3가지 접근방법 중에서 대부분의 연구들은 세 번째의 저차원 특징 기반 방법을 사용한다. 영상으로부터 손 영역을 검출 및 추적하고, 검출된 손 영역으로부터 위치나 움직임과 같은 특징들을 추출하여 사용한다.

최근 몇 년간 깊이 감지 기술의 발전과 함께, 실시간 깊이 카메라를 이용한 연구들이 진행되고 있다[2,3,4,5]. 특히, 최근에 출시된 마이크로소프트(Microsoft)사의 키넥트 센서는 저가의 깊이 카메라로써, 실시간으로 깊이 정보 뿐만 아니라 RGB 영상과 관절 추적 정보를 제공한다. 키넥트 센서로부터 제공되는 데이터(깊이, RGB, 관절 위치)의 사용은 제스처 인식을 위해 필요한 사람/신체부위 검출 및 포즈 추정의 수고를 덜어주고, 게임이나 인간-컴퓨터 상호작용 응용 개발을 쉽게 만들고 있다[6,7,8,9]. 따라서 본 논문에서는 키넥트에서 제공하는 관절 정보를 이용함으로써, 손 제스처와 관련된 신체 부위의 3차원 위치 정보를 쉽게 획득하여 손 제스처를 인식하는 데 사용한다.

손 제스처를 표현하기 위해 다양한 특징들이 사용되고 있다[4,5,6]. 손의 위치나 움직이는 속도 및 가속도는 오래전부터 사용된 특징이다. 그러나 위치 특징은 이동이나 회전 변화에 민감하다는 단점이 있으며, 속도 및 가속도 특징은 사람에 따라 다르게 추출될 수 있고 다양한 제스처 입력에 강인하지 못하다는 단점이 있다. 따라서 방향 특징이 많이 사용되고 있으며, 다양한 기준을 가지고 각도를 추출하여 방향 특징으로 사용하고 있다. 그러나 방향 특징 역시 제스처를 행하는 사람에 따라서 발생할 수 있는 방향 변형들을 효율적으로 표현하지 못하는 단점이 있다. 이러한 단점을 극복하기 위해, 본 논문에서는 손 제스처의 관찰열을 표현하기 위한 특징으로 방향 변형에 강인한 다각도 결합 히스토그램을 제안한다. 제안한 특징은 다양한 각도의 양자화 레벨을 갖는 여러 개의 각도 히스토그램들을 결합함으로써, 제스처를 행하는 사람에 따라 달라질 수 있는 방향 변형들을 허용하고 제스처 인식의 성능을 높인다. 또한, 제안한 특징은 제스처 관찰열의 클래스를 인식하기 위한 랜덤 결정 포레스트 분류기의 입력으로 사용되며, 분류기의 특성에 맞는 특징으로써 분류기와 잘 결합되어 높은 성능으로 제스처를 인식한다.

본 논문의 구성은 다음과 같다. Ⅱ장에서는 손 제스처 인식을 위한 제스처 표현 특징을 중심으로 관련 연구들에 대해 소개한다. Ⅲ장에서는 제안하는 손 제스처 인식 방법에 대해 자세히 설명하고, Ⅳ장에서는 다양한 실험 및 성능 평가를 통해 제안하는 방법의 효용성을 입증한다. 마지막으로 Ⅴ장에서는 결론 및 향후 연구 방향에 대해 기술한다.

Ⅱ. 관련 연구

손 제스처 인식을 위해 다양한 특징과 인식 방법이 사용되어 왔다. 대부분의 연구들은 손 제스처가 순차적인 데이터이기 때문에, 동적 시간 정합(Dynamic Time Warping, DTW) 알고리즘, 은닉 마르코프 모델(Hidden Markov Model, HMM) 또는 조건적 랜덤 필드(Conditional Random Field, CRF) 모델을 이용하여 제스처를 모델링하고 인식한다. 본 장에서는 인식 모델이 아닌, 제스처를 표현하는 특징들을 중심으로 관련 연구들을 분석한다.

손 제스처를 표현하기 위해 가장 많이 사용되는 특징은 손의 위치 정보이다[10,11]. Yang의 연구[10]에서는 미국인 수화(American Sign Language, ASL)를 인식하기 위해 움직임 정보로부터 손의 위치 특징을 추출하였고, 시간-지연 신경망(Time-delay neural network)을 이용하여 ASL 인식 시스템을 제안하였다. Doliotis의 연구[11]에서는 키넥트 센서로부터 획득된 깊이 정보를 이용하여 숫자를 나타내는 손 제스처를 인식하였다. 손 제스처를 인식하기 위한 특징은 손 영역의 정규화된 2D 위치를 사용하였고, DTW 알고리즘을 이용하여 손 제스처를 인식하였다.

위치 특징은 이동 변화나 회전에 민감하기 때문에, 방향 정보를 이용하여 손 제스처를 표현한 연구들도 있다[12,13]. Holden의 연구[12]에서는 오스트레일리아인(Australian) 수화를 인식하기 위해 머리를 기준으로 양손 사이 각도 및

손의 움직임 방향 등의 특징을 사용하였다. Yang의 연구[13]에서는 관절의 3차원 좌표를 각각 x, y, z축 평면으로 투영하고, 투영된 벡터와 등의 중심에서의 세로축과의 각도 정보를 특징으로 추출하였다.

위치와 방향 등의 특징을 결합하여 사용한 연구들도 많이 제안되었다[16,17,18,19]. [16,17,18]에서는 위치, 방향 및 속도 특징을 결합하여 손 제스처를 표현하는 데 사용하였고, Starner의 연구[19]에서는 손의 2D 위치, 프레임 간 손의 위치 변화, 축과의 각도, 손 영역의 고유벡터(eigenvector), 고유벡터의 크기 및 이심률(eccentricity) 특징들을 사용하였다.

제스처를 행하는 손 영역의 영상으로부터 제스처를 표현하는 특징을 추출하기도 한다[9,14,15]. Bergh의 연구[9]에서는 깊이 정보를 이용하여 검출된 손 영역 영상으로부터 Haarlet 계수(coefficient) 특징을 추출하였고, 최단 이웃(Nearest Neighbor, NN) 매칭을 통해 3차원 포인팅 제스처를 인식하였다. Ren의 연구[14,15]에서는 RGB 영상과 깊이 지도를 이용하여 손 형태을 검출하고, 손 형태는 시간-연속 커브 특징으로 표현하였다.

III. 제안하는 방법

1. 데이터

본 논문에서 손 제스처 인식을 위해 사용된 입력 데이터는 키넥트 센서로부터 획득된 사람의 관절 정보이다. 키넥트로부터 사람의 관절 정보를 획득하기 위해서는 사람의 대부분의 몸체가 키넥트 센서의 시야(Field of view, FOV) 내에 존재해야 한다. 일단, 대부분의 몸체가 키넥트 센서의 시야 내에 존재하게 되면, 키넥트 센서에서는 다음의 20개 관절에 대한 위치 정보를 제공한다: 머리, 왼쪽 어깨, 오른쪽 어깨, 어깨의 중간, 왼쪽 팔꿈치, 오른쪽 팔꿈치, 왼쪽 손목, 오른쪽 손목, 왼손, 오른손, 척추의 중앙, 왼쪽 둔부, 오른쪽 둔부, 둔부의 중간, 왼쪽 무릎, 오른쪽 무릎, 왼쪽 발목, 오른쪽 발목, 왼발, 오른발. 관절 위치는 키넥트의 깊이 센서를 중심으로 3차원 좌표를 미터(m) 단위로 표현하여 제공한다. 이렇게 획득된 20개의 관절 정보로부터, 손 제스처와 관련된 3개의 관절 정보(어깨, 팔꿈치, 손)를 사용하였다. 즉, 3개 관절에 대한 (x,y,z)좌표를 사용하였다. 그러나 제스처를 인식하기 위해 사용하는 관절의 선택은 제스처의 타입에 따라 달라질 수 있다.

2. 다각도 결합 히스토그램

키넥트 센서로부터 획득되는 관절에 대한 관측열을 X라고 할 때, X를 m개의 지역 관측들의 벡터 $X=\{x_1,x_2,\cdots,x_m\}$라고 하자. 1절에서 언급한 것처럼, 본 논문에서는 3개의 관절(어깨, 팔꿈치 및 손)의 움직임들로 구성된 손 제스처를 인식한다. 따라서 관측열 X의 각 원소 $x_i(i=1,2,\cdots,m)$는 다음의 식 (1)과 같이 3개의 관절에 대한 3차원 위치 벡터들로 구성된다.

$$x_i=\left[x_i^s,x_i^e,x_i^h\right],i=1,2,\cdots,m \tag{1}$$

위 식에서, x_i^s, x_i^e, x_i^h는 각각 어깨, 팔꿈치 및 손 관절에 대한 3차원 위치를 나타낸다.

사람의 관절 구조상, 팔꿈치는 어깨에 의해 관절의 3차원 위치가 결정되고, 손은 팔꿈치에 의해 관절의 3차원 위치가 결정된다. 이러한 손 제스처의 포즈를 표현하기 위해, 먼저 어깨를 기준으로 한 팔꿈치의 방향과 팔꿈치를 기준으로 한 손의 방향을 식 (2), (3)을 통해 계산한다.

$$\theta^e=\cos^{-1}\left(\frac{u_z}{\|u\|}\right),\phi^e=\tan^{-1}\left(\frac{u_y}{u_x}\right) \tag{2}$$

$$\theta^h=\cos^{-1}\left(\frac{v_z}{\|v\|}\right),\phi^h=\tan^{-1}\left(\frac{v_y}{v_x}\right) \tag{3}$$

위 식에서, u는 어깨와 팔꿈치 간 벡터이고, v는 팔꿈치와 손 간 벡터이다. u와 v벡터에 대한 구면(Spherical) 각도인 극각(Polar angle) θ와 방위각(Azimuthal angle) ϕ로 방향 특징을 추출한다 (그림 1 참고). 이렇게 추출된 방향 특징인 구면 각도는 손 제스처의 포즈를 표현하기 위한 특징 벡터로써 바로 사용될 수 있으나, 방향 변화에 강인하지 않다는 단점이 있다.

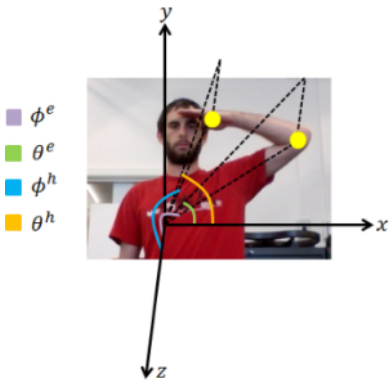

그림 1 . 구면 각도의 표현
Fig. 1. Graphical illustration of the spherical angles

이를 해결하기 위해 구면 각도로부터 각도 히스토그램을 계산한다. x_i에 대해 방향 특징으로 표현된 벡터를 $o_i=(\theta_i^e,\phi_i^e,\theta_i^h,\phi_i^h)$라 하면, 관측열 X의 모든 $o_i(i=1,2,\cdots,m)$를 이용하여 o_i 벡터의 각 요소에 대해 누적된 각도 히스토그램을 계산한다. θ_i^e와 θ_i^h에 대해서는 $0^\circ\sim180^\circ$ 사이의 각도를, ϕ_i^e와 ϕ_i^h에 대해서는 $-180^\circ\sim+180^\circ$ 사이의 각도를 일정한 간격으로 균등하게 분할하고, 관측열 X의 모든 m개의 지역 관측에 대해 히스토그램 빈(bin)에 누적한다. 이제 관측열 X는 γ각도 간격의 히스토그램 $h(\gamma)=(h^{\theta^e}(\gamma),h^{\phi^e}(\gamma),h^{\theta^h}(\gamma),h^{\phi^h}(\gamma))$로 표현된다. 추가적으로 방향 변형에 강인하게 하기 위해, 다른 각도의 양자화 레벨을 갖는 여러 개의 히스토그램을 결합한다. 이제 최종적으로 관측열 X는 식 (4)와 같이 표현된다.

$$H=(h(\gamma_1),h(\gamma_2),\cdots,h(\gamma_n)) \tag{4}$$

$$h(\gamma_j)=\left(h^{\theta^e}(\gamma_j),h^{\phi^e}(\gamma_j),h^{\theta^h}(\gamma_j),h^{\phi^h}(\gamma_j)\right),\ j=1,2,\cdots,n \tag{5}$$

위 식에서, n은 최종 히스토그램 H를 구성하는 데 사용된 각도 히스토그램의 개수를 나타낸다. H의 각 요소는 서로 다른 γ_j각도 간격의 히스토그램으로써, 방향 변형에 강인하게 동작한다. 그림 2는 제안하는 다각도 결합 히스토그램을 생성하는 과정을 보여준다.

3. 제스처 인식

2절에서 제안한 다각도 결합 히스토그램에 의해 표현된 제스처를 인식하기 위해 랜덤 결정 포레스트(Randomized decision forests) 분류기를 사용한다. 랜덤 결정 포레스트[20]는 무작위로 특징들의 부분집합을 선택하여 분류기를 학습시키는 방식으로 작동하며, 여러 개의 결정 트리들로 구성된 앙상블(ensemble) 분류기이다. 속도가 빠르고 큰 데이터 셋에서도 효율적으로 동작하며, 다른 다중 클래스 분류기에 비해 높은 정확도를 준다는 것이 입증되어왔다. 랜덤 결정 포레스트 분류기의 학습과 테스트 과정은 다음과 같다.

1.1 학습

T개의 결정 트리들로 구성된 랜덤 포레스트는 학습 히스토그램 특징 데이터 H로부터 교체(Replacement)방식으로 무작위로 선택된 부분집합 H'을 이용하여 생성된다. 각 결정 트리는 랜덤 부분집합 H'을 분할(Split) 함수와 임계값(Threshold)을 이용하여 반복적으로 왼쪽 부분집합 H_L과 오른쪽 부분집합 H_R으로 분할함으로써 학습한다. 최종적으로, 랜덤 포레스트 분류기는 가장 큰 엔트로피 차이(Information gain)를 갖는 모든 부분집합들의 분할 조건(특징 파라미터, 임계값, 분할 함수) 집합을 찾음으로써 생성된다.

1.2 테스트

다각도 결합 히스토그램 특징으로 표현된 테스트 제스처

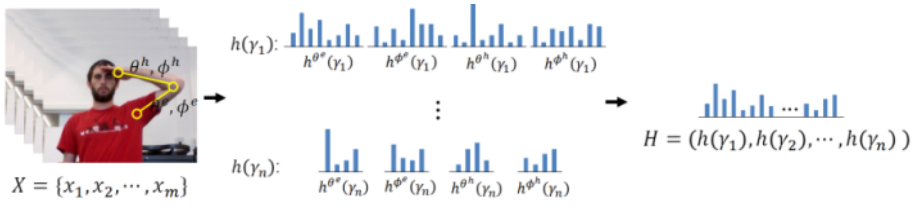

그림 2 . 다각도 결합 히스토그램 생성 과정
Fig. 2. Graphical illustration of the combined angle histogram

의 클래스를 인식하기 위해, 학습된 랜덤 포레스트에 테스트 데이터 H_S를 입력값으로 넣는다. 그 결과, 랜덤 포레스트의 각 결정 트리 t의 단말(Leaf) 노드에는 각 제스처 클래스 c에 대한 확률 분포 $P_t(c|H_S)$가 저장된다. 최종 클래스 확률 분포는 랜덤 포레스트에 있는 모든 트리들에 대한 확률 분포들의 평균으로 계산하며, 최종 제스처 클래스 $\hat{c}$은 다음의 식 (6)을 통해 결정된다.

$$\hat{c} = \underset{c}{\arg max}\left\{\frac{1}{T}\sum_{t=1}^{T} P_t(c|H_S)\right\} \quad (6)$$

Ⅳ. 실험

1. 데이터셋

제안하는 방법의 성능을 평가하기 위해, 키넥트 센서

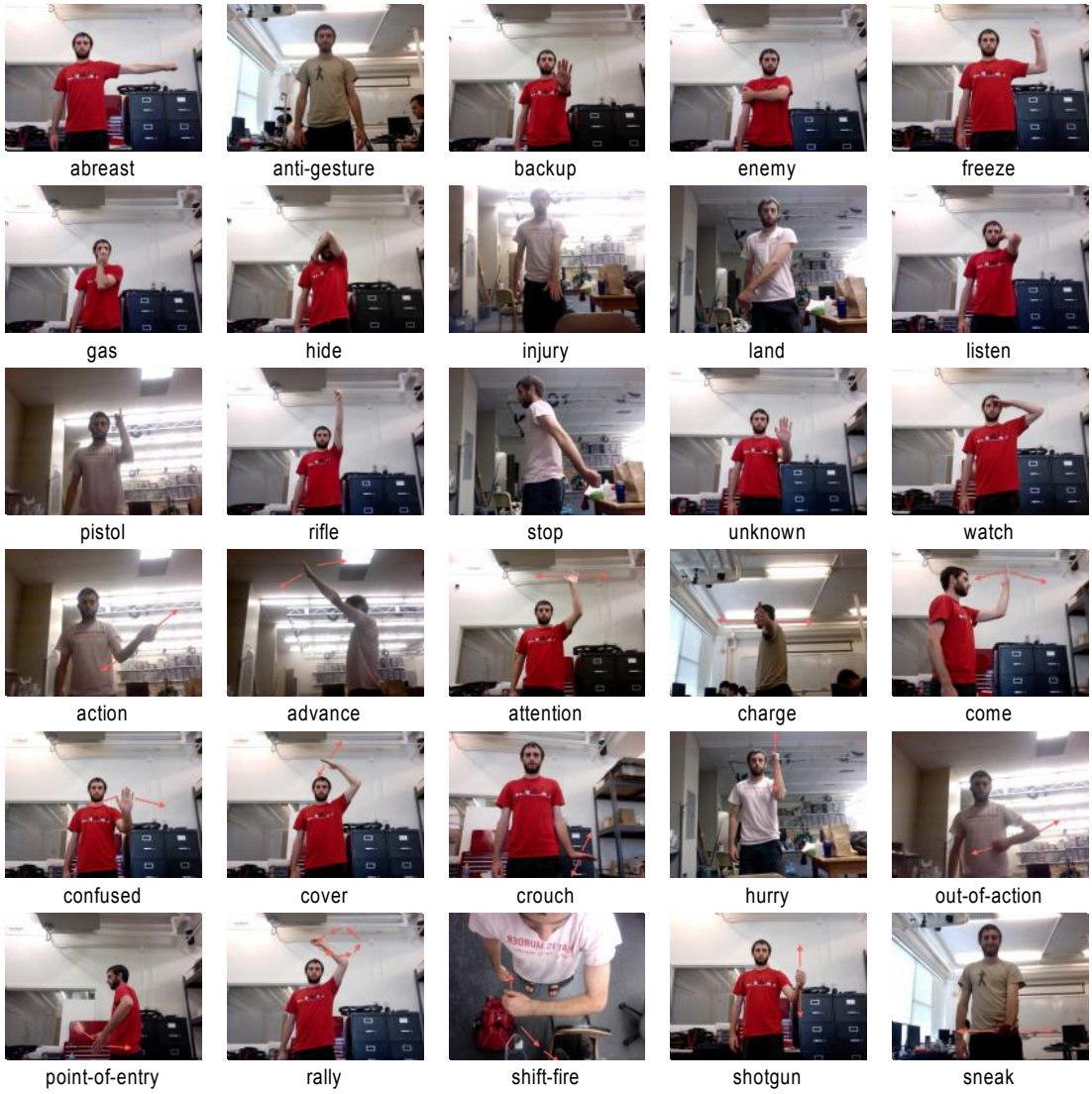

그림 3. Cornell 손 제스처 데이터셋의 예
Fig. 3. Example snapshots of the Cornell hand gesture dataset [21]

를 이용하여 촬영한 Cornell 손제스처 데이터셋[21]과 MSRDailyActivity3D[22]을 사용하였다. Cornell 손제스처 데이터셋은 3명의 다른 사람에 의해 촬영된 900개의 비디오 클립으로 구성되어 있으며, 각 비디오 시퀀스는 사람의 어깨, 팔꿈치, 손에 대한 관절 정보로 이루어져 있다. 군사 및 비행 제어를 위해 사용되는 30개 클래스의 손제스처 데이터로 구성되어 있으며, 15개의 정적 타입과 15개의 동적 타입의 손제스처를 포함한다. 각 비디오 시퀀스는 초당 50 프레임으로 5초간 촬영되었다. 그림 3은 정적 타입과 동적 타입에 대한 손제스처 데이터를 보여준다. 학습을 위해 각 제스처 클래스 당 15개씩 총 450개의 비디오 클립을 사용하였고, 나머지는 테스트를 위해 사용하였다.

MSRDailyActivity3D 데이터셋은 총 16개 타입의 행동 클래스로 구성되어 있으며, 10명의 다른 사람이 각 행동을 앉아있거나 서있는 상태로 수행한다. 각 행동 클래스 당 20개씩 총 320개의 비디오 클립(2~10초), 깊이지도 및 관절 정보를 포함하고 있다. 학습 및 테스트를 위해 각각 160개의 비디오 클립에 대한 관절 정보를 사용하였다. 이 데이터셋을 사용한 이유는 제안하는 방법이 손 제스처 뿐만 아니라 행동 인식까지 확장될 수 있다는 것을 보이기 위해 사용하였다.

2. 실험 결과 및 분석

제안하는 다각도 결합 히스토그램 특징에 대한 효용성을 입증하기 위해, 각도 특징에 기반한 방법과 제안하는

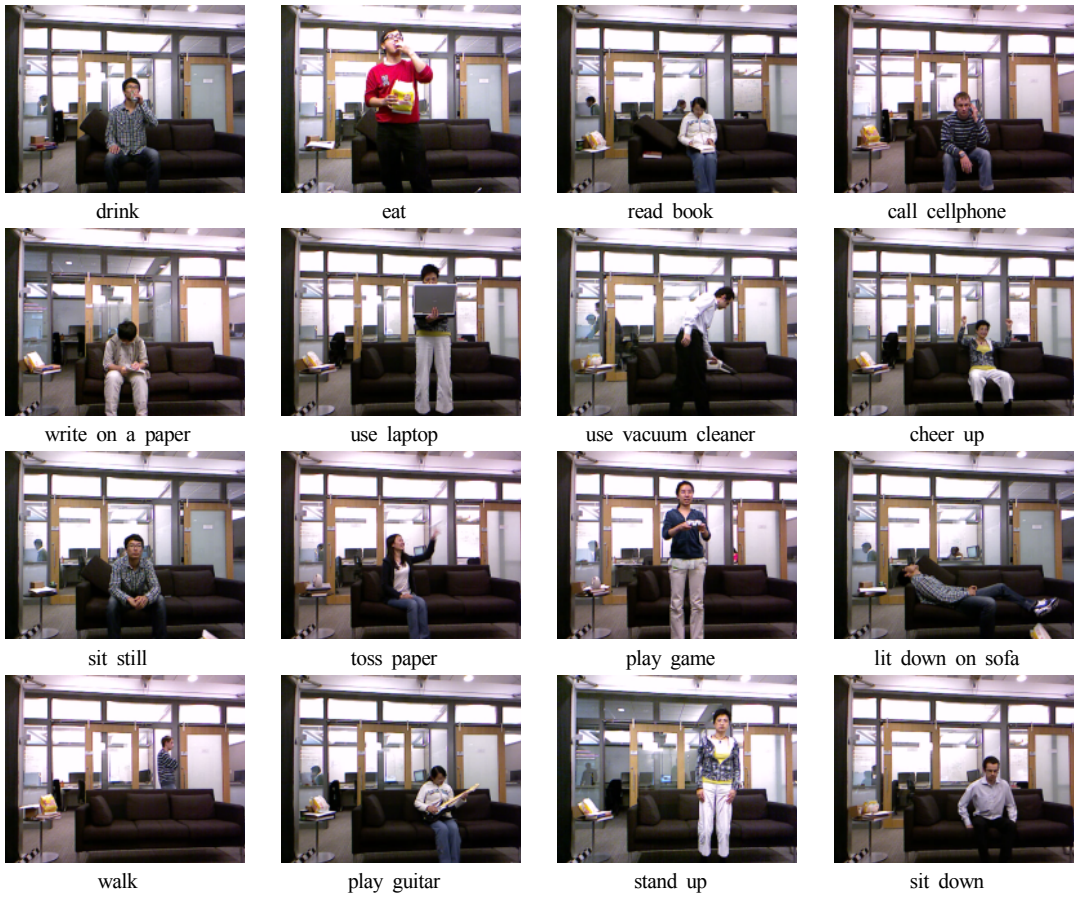

그림 4. MSRDailyActivity3D 데이터셋의 예
Fig. 4. Example snapshots of the MSRDailyActivity3D dataset [22]

다각도 결합 히스토그램 특징에 기반한 방법간의 인식 성능을 비교하였다. 각도 특징에 기반한 방법은 구면 각도 특징과 랜덤 포레스트 분류기를 사용하였다. Cornell 손제스처 데이터셋에 대해서는 앞에서 언급한 것처럼 어깨, 팔꿈치, 손의 관절 정보로부터 추출한 특징을 사용하였고, MSRDailyActivity3D 데이터셋에 대해서는 사람의 행동을 표현하기 위해 어깨, 팔꿈치, 손의 관절로부터 추출되는 특징 뿐만 아니라, 둔부, 무릎, 발의 관절 정보로부터 추출한 특징까지 사용하였다.

그림 5는 랜덤 포레스트 분류기내 결정 트리 개수에 따른 두 방법 간 인식 비교 성능을 보여준다. 제안하는 방법은 다른 개수의 결정 트리를 갖는 모든 랜덤 포레스트 분류기에서 각도 특징 기반 방법보다 높은 인식 성능을 획득하였다. 1, 20, 40, 60, 80, 100개의 트리를 가진 랜덤 포레스트 분류기에 대해, Cornell 손제스처 데이터셋에서는 3.33%, 7.34%, 5.11%, 5.34%, 4.89%, 4.67%의 인식 성능 향상이 있었고, 평균적으로 5.86%의 성능 향상이 있었다. MSR DailyActivity3D 데이터셋에서는 9.37%, 11.25%, 8.75%, 12.5%, 16.25%, 8.75%의 인식 성능 향상이 있었고, 평균적으로 11.8%의 성능 향상이 있었다.

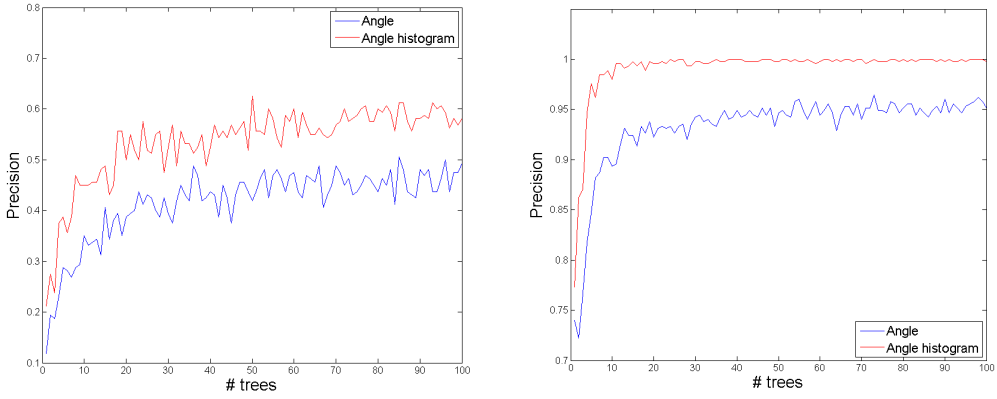

그림 5. 랜덤 포레스트 분류기 내 결정 트리 개수에 따른 각도 특징 기반과 제안하는 방법간의 인식 성능 비교. (a) Cornell 손 제스처 데이터셋, (b) MSRDailyActivity3D 데이터셋
Fig. 5. The comparison of recognition accuracy between our combined angle histogram and angle-based approach. (a) Cornell hand gesture dataset, (b) MSR Daily Activity 3D dataset

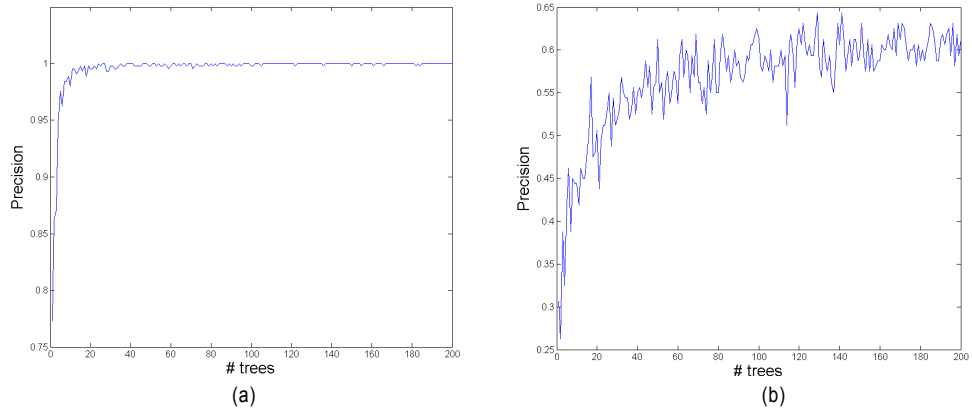

그림 6. 랜덤 포레스트 분류기 내 결정 트리 개수에 따른 평균 인식 정확도. (a) Cornell 손 제스처 데이터셋, (b) MSRDailyActivity3D 데이터셋
Fig. 6. The average recognition accuracy according to the different number of trees in the forest. (a) Cornell hand gesture dataset, (b) MSRDailyActivity3D dataset

그림 6은 랜덤 포레스트 분류기 내 결정 트리 개수에 따른 평균 인식 정확도 성능을 보여준다. 1개부터 200개까지 결정 트리의 개수를 1개씩 늘리면서 성능을 측정하였고, 10°, 20°, 30° 의 각도 간격을 갖는 3개의 각도 히스토그램들을 결합하여 특징으로 사용하였다. 평균적으로 결정 트리의 개수가 증가할 때마다 더 높은 인식 성능을 얻었으며, Cornell 손제스처 데이터셋에서는 85개 이상의 결정 트리를 갖는 랜덤 포레스트 분류기부터 100%의 인식 정확도를 갖는 경우가 많았다. MSRDailyActivity3D 데이터셋에서는 100개 이상의 결정 트리를 갖는 랜덤 포레스트 분류기의 평균 인식 정확도가 59.91%였고, 특히 결정 트리의 개수가 141개 일 때, 64.38%의 가장 높은 정확도를 획득하였다.

단일 각도 히스토그램 특징 기반 방법과 제안하는 다각도 결합 히스토그램 특징 기반 방법간의 인식 성능을 비교하였다. 각도 히스토그램 특징에 대해, 각각 10°, 20°, 30° 의 각도 간격을 갖는 히스토그램을 사용하였다. 3개의 각도를 선택한 이유는, 20° 의 각도 간격을 갖는 히스토그램 특징이 가장 높은 성능을 가졌기 때문에 20° 주변의 3개 각도 간격을 갖는 히스토그램에 대해서만 성능을 비교한 것이다. 다각도 결합 히스토그램은 3개의 각도 (10°, 20°,30°) 히스토그램들을 결합하였다. 표 1과 2는 다른 결정 트리 개수에 따른 두 방법 간 성능 비교 결과를 보여주는데, 다음의 3가지 결론을 도출할 수 있다. 첫 번째, 각도 히스토그램 특징을 이용한 방법이 각도 특징을 이용한 방법보다 높은 인식 성능을 갖는다. 그림 5의 각도 특징 기반 방법의 인식 성능에서 볼 수 있듯이, 다른 결정 트리 개수를 갖는 모든 랜덤 포레스트 분류기에서 각도 히스토그램 특징을 이용한 방법의 인식 성능이 높게 나왔다. 이는 본 논문에서 사용한 히스토그램 표현의 효용성을 보여주는 것이다. 두 번째, 3개 각도 간격의 히스토그램 특징을 이용한 방법 중에서, 20° 각도의 히스토그램 특징이 대부분의 결정 트리 개수에서 가장 좋은 성능을 획득한다. 세 번째, 제안하는 다각도 결합 히스토그램 특징은 단일 각도 히스토그램 특징보다 2개 이상의 다중 결정 트리를 갖는 분류기에 대해 항상 높은 성능을 갖는다.

표 1. Cornell 손 제스처 데이터셋에서의 각도 히스토그램과 제안하는 다각도 결합 히스토그램 특징 간 인식 성능 비교
Table 1. The comparison of recognition accuracy between combined angle histogram and single angle histogram approaches on Cornell hand gesture dataset

결정 트리 개수	10°	20°	30°	결합
1	0.8244	0.5533	0.6667	0.8044
4	0.8889	0.9000	0.8378	0.9156
6	0.9244	0.9311	0.9111	0.9644
10	0.9711	0.9844	0.9622	0.9889
30	0.9822	0.9956	0.9933	0.9956
100	0.9956	0.9978	0.9956	1.0000

표 2. MSRDailyActivity3D 데이터셋에서의 각도 히스토그램과 제안하는 다각도 결합 히스토그램 특징 간 인식 성능 비교
Table 2. The comparison of recognition accuracy between combined angle histogram and single angle histogram approaches on MSRDailyActivity3D dataset

결정 트리 개수	10°	20°	30°	결합
1	0.2000	0.2563	0.2500	0.3063
7	0.3375	0.3188	0.3500	0.3875
10	0.4000	0.3250	0.4063	0.4187
30	0.4563	0.4750	0.4625	0.5188
50	0.5000	0.5625	0.5188	0.6250
100	0.5563	0.5500	0.4938	0.6125

표 3은 50개의 결정 트리를 갖는 랜덤 포레스트 분류기에 대해 다양한 각도 히스토그램의 조합에 따른 인식 성능을 보여준다. Cornell 손제스처 데이터셋에서는 $(10^\circ, 20^\circ, 30^\circ)$와 $(15^\circ, 20^\circ, 30^\circ)$ 조합에서의 인식 성능이 가장 높았으며, MSRDailyActivity3D 데이터셋에서는

표 3. 다양한 각도 히스토그램의 조합에 대한 인식 성능 비교
Table 3. The comparison of recognition accuracy according to various combinations of angle histograms

조합 방법	Cornell 손제스처 데이터셋 인식 정확도	MSRDailyActivity3D 데이터셋 인식 정확도
$(5^\circ, 10^\circ, 15^\circ)$	0.9933	0.5500
$(10^\circ, 20^\circ, 30^\circ)$	1.0000	0.6250
$(15^\circ, 20^\circ, 30^\circ)$	1.0000	0.6062
$(20^\circ, 40^\circ, 60^\circ)$	0.9978	0.6062
$(30^\circ, 40^\circ, 60^\circ)$	0.9956	0.5750

$(10^\circ, 20^\circ, 30^\circ)$ 조합에서의 인식 성능이 가장 높았다. 이는 본 절의 제안하는 방법의 특징으로 사용된 결합 히스토그램이 $(10^\circ, 20^\circ, 30^\circ)$의 조합을 이용한 것의 타당성을 보여주는 것이다.

표 4와 5는 Cornell 손제스처 데이터셋에서 정적 및 동적 타입의 각 제스처에 대해 다양한 방법과의 성능을 비교하여 보여준다. 첫 번째 방법[21]은 관절 각도들의 사인곡선적(Sinusoidal) 빈도, 진폭, 평균 및 표준편차를 손 제스처의 특징으로 사용하였다. 제스처를 분류하기 위해서는 다중 클래스 서포트 벡터 머신(Multi-class Support Vector Machine, MSVM) 분류기가 사용되었고, 0.3의 c값과 RBF 커널 함수가 사용되었다. 두 번째 방법은 본 논문에서 제안한 다각도 결합 히스토그램을 특징으로 사용하였고, 제스처 분류를 위해서는 [21]과 같은 파라미터를 갖는 MSVM 분류기를 사용하였다. 제안하는 방법은 다각도 결합 히스토그램 특징과 100개의 결정 트리를 갖는 랜덤 포레스트 분류기를 사용한 것의 결과이다. 먼저, [21]과 두 번째 방법을 비교해 보면, 두 방법의 차이는 같은 MSVM 분류기를 사용하였지만 서로 다른 특징을 사용한 것이다. 표 4를 보면 정적 타입의 제스처에 대해 두 번째 방법의 성능이 더 높으며, 표 5의 동적 타입의 제스처에 대해서는 첫 번째 방법의 성능이 약간 더 높지만 거의 차이가 없는 것을 볼 수 있다. 이는 본 논문에서 제안하는 다각도 결합 히스토그램 특징의 효용성을 보여주는 것이다. 두 번째 방법과 제안하는 방법을 비교해 보면, 두 방법의 차이는 같은 특징을 사용하였지만 서로 다른 분류기를 사용한 것이다. 표에서 볼 수 있듯이, 제안하는 방법은 정적 및 동적 타입의 모든 제스처에 대해 100%의 높은 인식 정확도를 획득하였다. 이로부터 제안하는 다각도 결합 히스토그램 특징이 랜덤 포레스트 분류기와 함께 잘 동작하여 높은 인식 성능을 제공한다는 것을 알 수 있다. 따라서 이러한 실험 결과는 본 논문에서 제안한 손 제스처를 인식하기 위한 특징과 분류기가 효율적이라는 사실을 보여준다.

표 4. Cornell 손제스처 데이터셋의 정적 타입의 제스처에서의 [21]의 방법과 제안하는 방법의 인식 성능 비교
Table 4. Comparison with [21] on the static gestures of Cornell hand gesture dataset

정적 제스처 종류	[21]	제안 특징+MSVM	제안하는 방법
abreast	1.00	1.00	1.00
anti-gesture	0.99	1.00	1.00
backup	0.95	1.00	1.00
enemy	1.00	1.00	1.00
freeze	1.00	1.00	1.00
gas	0.90	1.00	1.00
hide	1.00	1.00	1.00
injury	1.00	1.00	1.00
land	0.93	1.00	1.00
listen	0.74	1.00	1.00
pistol	1.00	1.00	1.00
rifle	1.00	1.00	1.00
stop	1.00	1.00	1.00
unknown	1.00	0.93	1.00
watch	0.85	1.00	1.00

표 5. Cornell 손제스처 데이터셋의 동적 타입의 제스처에서의 [21]의 방법과 제안하는 방법의 인식 성능 비교
Table 5. Comparison with [21] on the dynamic gestures of Cornell hand gesture dataset

동적 제스처 종류	[21]	제안 특징+MSVM	제안하는 방법
action	1.00	1.00	1.00
advance	1.00	1.00	1.00
attention	1.00	0.87	1.00
charge	1.00	1.00	1.00
come	0.91	1.00	1.00
confused	0.98	1.00	1.00
cover	1.00	1.00	1.00
crouch	1.00	1.00	1.00
hurry	0.98	1.00	1.00
out-of-action	0.96	1.00	1.00
point-of-entry	0.99	1.00	1.00
rally	1.00	0.87	1.00
shift-fire	1.00	1.00	1.00
shotgun	0.98	1.00	1.00
sneak	0.98	1.00	1.00

표 6은 MSRDailyActivity3D 데이터셋에서의 제안하는 다각도 결합 히스토그램 특징 및 MSVM 분류기를 사용한 방법과 제안하는 방법(다각도 결합 히스토그램 특징+랜덤 포레스트 분류기)간의 성능을 비교하여 보여준다. MSVM 분류기를 사용한 방법은 평균적으로 35.62%의 정확도를,

제안하는 방법은 89개의 결정 트리를 갖는 랜덤 포레스트 분류기를 사용했을 때 63.75%의 정확도를 획득하였다. 대부분의 행동 클래스에 대해 제안하는 방법이 더 높은 정확도를 획득하였으며, 이는 제안하는 특징이 랜덤 포레스트 분류기와 함께 잘 동작한다는 것을 나타낸다.

표 6. MSRDailyActivity3D dataset 에서의 다른 분류기 기반 방법 간 인식 성능 비교
Table 6. Comparison between methods based on different classifiers on MSRDailyActivity3D dataset

행동 클래스	제안 특징+MSVM	제안하는 방법
drink	0.30	0.90
eat	0.90	0.70
read book	0.60	0.60
call cellphone	0.20	0.10
write on a paper	0.00	0.30
use laptop	0.10	0.30
use vacuum cleaner	0.70	0.90
cheer up	0.50	1.00
sit still	0.20	0.80
toss paper	0.00	0.60
play game	0.20	0.30
lie down on sofa	0.40	0.60
walk	0.20	0.60
play guitar	0.20	0.70
stand up	0.90	1.00
sit down	0.30	0.80

그림 7은 MSRDailyActivity3D 데이터셋에서의 결정 트리 개수가 증가함에 따른 학습 및 테스트 시간의 변화를 보여준다. 행동을 인식하기 위해 어깨, 팔꿈치, 손의 관절로부터 추출한 특징을 이용한 방법(4개 히스토그램 결합: 4H)과 힙, 무릎, 발의 관절까지 포함하여 추출한 특징을 이용한 방법(8개 히스토그램 결합: 8H)을 비교하였다. 테스트 데이터(총 160개 비디오 클립)의 특징을 추출하는데 4H 방법에서는 총 0.5229초, 8H 방법에서는 총 0.9584초가 걸렸다. 그림 7은 특징을 추출한 후, 각 결정 트리 개수에 따른 학습 및 테스트 시간을 나타낸 것이다. 학습 시간은 트리 개수가 증가함에 따라 선형적으로 증가함을 확인하였고, 테스트 시간은 트리 개수가 증가함에도 불구하고 일정함을 확인하였다. 또한, 다른 특징 차원에 대해서도 테스트 시간은 일정함을 확인하였다. 이는 제안하는 방법이 특징 차원 및 트리 개수 증가에 따라 시간 복잡도가 증가하지 않음을 의미한다. 그림 6에서 나타낸 것처럼, 제안하는 방법은 트리 개수가 증가함에 따라 인식 정확도가 높아지는 특성을 가지고 있기 때문에, 시간 복잡도 고려 없이 높은 인식률을 획득하는 트리 개수로 정해주면 된다.

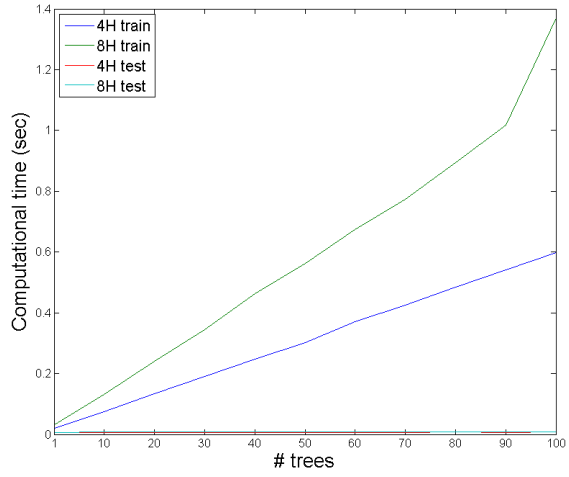

그림 7. 다른 결정 트리 개수 및 특징 차원에 따른 계산 시간
Fig. 7. The computational time according to the different number of trees and feature dimensions

Ⅴ. 결론

본 논문에서는 키넥트 센서로부터 획득한 관절 정보를 이용하여 손 제스처를 인식하는 방법을 제안하였다. 손 제스처를 표현하기 위한 특징으로써 다양한 관절 각도의 움직임을 표현할 수 있는 다각도 결합 히스토그램을 제안하였다. 제안한 특징은 다양한 각도의 양자화 레벨을 갖는 여러 개의 각도 히스토그램들을 결합함으로써 방향 변화에 강인하도록 하였다. 또한 제스처를 인식하기 위해 랜덤 포레스트 분류기를 사용하였으며, 실험을 통해 제안한 특징과 함께 사용된 랜덤 포레스트 분류기의 성능이 다른 방법에 비해 우수함을 보였다. 향후 연구로는, 관절 정보 뿐만 아니라 키넥트 센서로부터 획득 가능한 RGB 영상 또는 깊

이 데이터까지 함께 활용하여 손 제스처 인식의 성능을 향상시키는 것이다.

참 고 문 헌

[1] G.R.S. Murthy and R.S. Jadon, "A review of vision based hand gestures recognition," Journal of Information Technology and Knowledge Management, vol. 2, no. 2, pp. 405-410, 2009.

[2] V. Ganapathi, C. Plageman, D. Koller, and S. Thrun, "Real time motion capture using a single time-of-flight camera," In Conf. on Computer Vision and Pattern Recognition, pp. 755-762, 2010.

[3] M. Siddiqui and G. Medioni, "Human pose estimation from a single view point, real-time range sensor," In Workshop on Computer Vision for Computer Games at Conf. on Computer Vision and Pattern Recognition, pp. 1-8, 2010.

[4] R. Munoz-Salinas, R. Medina-Carnicer, F.J. Madrid-Cuevas, and A. Carmona-Poyato, "Depth silhouettes for gesture recognition," Pattern Recognition Letters, vol. 29, no. 3, pp. 319-329, 2008.

[5] P. Suryanarayan, A. Subramanian, and D. Mandalapu, "Dynamic hand pose recognition using depth data," In Conf. on Pattern Recognition, pp. 3105-3108, 2010.

[6] I. Oikonomidis, N. Kyriazis, and A.A. Argyros, "Efficient model-based 3D tracking of hand articulations using Kinect," In British Machine Vision Conference, pp. 101.1-101.11, 2011.

[7] J. Sung, C. Ponce, B. Selman, and A. Saxena, "Human activity detection from RGBD images," In AAAI 2011 Workshop, pp. 47-55, 2011.

[8] W. Li, Z. Zhang, and Z. Liu, "Action recognition based on a bag of 3d points," In Computer Vision and Pattern Recognition Workshops(CVPRW), pp. 9-14, 2010.

[9] M.V. Bergh, D. Carton, R.D. Nijs, N. Mitsou, C. Landsiedel, K. Kuehnlenz, D. Wollherr, L.V. Gool, and M. Buss, "Real-time 3D hand gesture interaction with a robot for understanding directions from humans," In Symposium on Robot and Human Interactive Communication, pp. 357-362, 2011.

[10] M. Yang, N. Ahuja, and M. Tabb, "Extraction of 2D motion trajectories and its application to hand gesture recognition," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, no. 8, pp. 1061-1074, 2002.

[11] P. Doliotis, A. Stefan, C. McMurrough, D. Eckhard, and V. Athitsos, "Comparing gesture recognition accuracy using color and depth information," PETRA, pp. 1-7, 2011.

[12] E.-J. Holden, G. Lee, and R. Owens, "Australian sign language recognition," Machine Vision and Applications, vol. 16, no. 5, pp. 312-320, 2005.

[13] H.-D. Yang, A.-Y. Park, and S.-W. Lee, "Gesture spotting and recognition for human-robot interaction," IEEE Trans. on Robotics, vol. 23, no. 2, pp. 256-270, 2007.

[14] Z. Ren, J. Yuan, and Z. Zhang, “Robust hand gesture recognition with kinect sensor," Proc. of ACM Intl. Conf. on Multimedia, pp. 759-760, 2011.

[15] Z. Ren, J. Meng, and J. Yuan, "Depth camera based hand gesture recognition and its applications in human-computer interaction," IEEE International Conference on Information, Communication, and Signal Processing, pp. 1-5, 2011.

[16] Y. Ho-Sub, S. Jung, J.B. Young, and S.Y. Hyun, "Hand gesture recognition using combined features of location, angle and velocity," Journal of Pattern Recognition, vol. 34, no. 7, pp. 1491-1501, 2001.

[17] M. Elmezain, A. Al-Hamadi, and B. Michaelis, "Improving hand gesture recognition using 3D combined features," International Conference on Machine Vision, pp. 128-132, 2009.

[18] H.-S. Yoon, J. Soh, Y.J. Bae, and H.S. Yang, "Hand gesture recognition using combined features of location, angle and velocity," Pattern Recognition, vol. 32, no. 7, pp. 1491-1501, 2001.

[19] T. Starner, J. Weaver, and A. Pentland, "Real-time american sign language recognition using desk and wearable computer based videos," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 20, no. 12, pp. 1371-1375, 1998.

[20] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.

[21] http://pr.cs.cornell.edu/humanactivities/handgesturedata.html

[22] http://research.microsoft.com/en-us/um/people/zliu/actionrecorsrc/default.htm

저 자 소 개

조 선 영

- 2007년 : 숙명여자대학교 컴퓨터과학과 졸업(이학사)
- 2009년 : 연세대학교 컴퓨터과학과 석사 졸업(공학석사)
- 2009년 ~ 현재 : 연세대학교 컴퓨터과학과 박사과정
- 주관심분야 : 제스처 인식, 사람 행동 인식, 패턴 인식

변 혜 란

- 1980년 : 연세대학교 수학과 졸업(이학사)
- 1983년 : 연세대학교 수학과 졸업(이학석사)
- 1987년 : University of Illinois, Computer Science(M.S.)
- 1993년 : Purdue University, Computer Science(Ph.D.)
- 1994년 ~ 1995년 : 한림대학교 정보공학과 조교수
- 1995년 ~ 현재 : 연세대학교 컴퓨터과학과 교수
- 주관심분야 : 패턴 인식, 영상 처리, 영상 인식

이 희 경

- 1999년 : 영남대학교 컴퓨터공학과 졸업
- 2002년 : KAIST-ICC 정보통신공학부 졸업
- 2002년 ~ 현재 : 한국전자통신연구원 방통융합미디어연구부 선임연구원
- 주관심분야 : HCI, Gaze Tracking, Bi-directional AD, 맞춤형방송

차 지 훈

- 1992년 : 명지대학교 전자계산공학사
- 1996년 : Florida Institute of Technology 공학석사 (Computer Science)
- 2002년 : Florida Institute of Technology 공학박사 (Computer Science)
- 2003년 ~ 현재 : 한국전자통신연구원 방통융합미디어연구부 융합미디어연구팀장
- 2008년 3월 ~ 현재 : 과학기술연합대학원대학교 부교수
- 주관심분야 : Multimedia streaming, Interactive broadcasting system, Feature extraction/tracking, Richmedia