키넥트 센서를 이용한 팔 제스처 인식 시스템의 설계

허세경, 신예슬, 김혜숙, 김인철
경기대학교 컴퓨터과학과
e-mail : {tprudzzang, ysulshin, chia, kic}@kyonggi.ac.kr

Design of an Arm Gesture Recognition System using Kinect Sensor

Se-Kyeong Heo, Ye-Seul Shin, Hye-Suk Kim, In-Cheol Kim
Dept of Computer Science, Kyonggi University

요 약

최근 카메라 영상을 이용한 제스처 인식 관련 연구가 활발히 진행되고 있다. 카메라 영상을 이용한 제스처 인식에서 많이 사용되는 학습 알고리즘에는 확률 그래프 모델인 HMM과 CRF 등이 있다. 이 학습 알고리즘들은 다차원의 연속된 실수 데이터를 가지고 모델을 학습하면 계산량이 많아진다. 본 논문에서는 팔 관절 위치 데이터를 k-평균 군집화 과정을 거쳐 1차원의 시계열 데이터로 변환 후, 제스처별로 HMM 모델을 학습하는 방법을 제안한다. 키넥트 센서를 통해 얻은 팔 관절 위치 데이터에 k-평균 군집화를 적용하여 1차원 시계열 데이터를 생성하고, 이를 HMM의 학습 및 인식에 사용한다. 본 논문에서 제안하는 방법의 성능을 분석하기 위하여, 다른 시계열 학습 알고리즘인 AP+DTW를 이용한 방법과의 비교 실험을 포함해 다양한 실험들을 수행하였다.

1. 서론

기계와 상호작용을 하기위한 현재 연구되는 기술 중에서 인간이 실생활에 가장 많이 사용하는 신체 부위 중 하나인 팔을 이용한 제스처 인식이 각광 받고 있다. 팔을 이용한 제스처 인식을 위해서는 기계가 인식할 수 있는 팔의 위치를 알아야 한다. 팔의 위치를 알아내기 위한 센서는 2D 카메라, RF-ID 센서 등이 있으며, 특히 2010년 말 등장한 마이크로소프트(Microsoft)의 키넥트(Kinect) 센서는 실생활에 사용 가능한 정확도와 실시간 관절 추적 시스템, RGB 영상 등의 다양한 기능을 저가로 지원한다[1]. 키넥트 센서를 이용하면 제스처 인식을 위한 팔의 관절 위치 데이터 또는 좌표 데이터를 신체 부위 검출이나 포즈 추정 과정 없이 바로 얻을 수 있다[2,3]. 제스처 인식에서 많이 사용되는 학습 알고리즘에는 확률 그래프 모델인 HMM(Hidden Markov Model)과 CRF(Conditional Random Field) 등이 있다. 이 학습 알고리즘들은 다차원의 연속된 실수 데이터를 가지고 모델을 학습하면 계산량이 많아진다.

본 논문에서는 팔 제스처 인식을 위해 키넥트 센서를 이용하여 팔 제스처 인식을 위한 좌표 데이터를 수집한다. 단, 좌표 데이터는 위치를 나타내는 공간적인 좌표이므로 이동과 회전에 민감하다. 이를 해결하기 위해 좌표 데이터를 각도 데이터로 변환하여 사용한다. 그리고 데이터의 전처리 과정에서 각도 데이터를 k-평균 군집화(k-Means Clustering)를 통해 유사한 데이터를 군집화 한다. 그 결과 8차원의 각도 데이터를 1차원의 정수 데이터로 변환한다. 이 과정은 학습 알고리즘이 제스처 종류별로 모델을 학습하기 위한 계산량을 감소시켜, 팔 제스처의 데이터를 실시간으로 분석하도록 한다. 본 논문에서는 1차원의 정수 시계열 데이터를 이용하여 제스처별 HMM 모델을 학습 후 제스처를 인식하고자 한다. 그리고 제안하는 방법을 평가하기 위해, 다른 시계열 학습 알고리즘을 이용해 모델을 학습하고 두 학습 방법을 비교 실험하고자 한다.

※ 본 연구는 경기도의 경기도지역협력연구센터사업의 일환으로 수행하였음

2. 관련연구

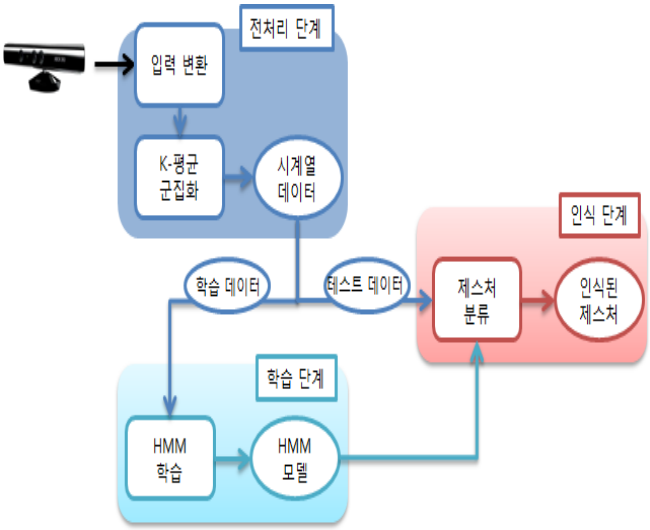
Lee의 연구[4]에서는 카메라 영상을 통한 손 제스처 인식 기능을 이용해, 파워포인트(PowerPoint)의 슬라이드 쇼를 제어하는 연구를 진행하였다. 연속적인 제스처 입력으로부터 의미 있는 동작인 제스처와 비 제스처를 구분하고자, HMM 알고리즘을 기반으로 하는 적응적 임계치 모델을 제안하였다. 적응적 임계치 모델을 사용하면 기존의 임계치 모델을 했을 때 임계치를 결정해야 하는 문제를 해결할 수 있다. 그러나 적응적 임계치 모델의 상태 수가 제스처 집합의 상태수를 모두 합한 값이므로, 구분해야 하는 제스처가 많아질수록 HMM 학습 계산량이 많다는 단점이 있다. Cho의 연구[5]에서는 팔 제스처 인식을 위해 2 계층 제스처 추출 인식 모델을 제안하였다. 키넥트 센서를 이용해 받은 연속적인 데이터를 k-평균 군집화를 통해 특징 벡터로 군집화 한다. 계층1에서는 프레임마다 HMM 확률 값을 적용하여 제스처와 비 제스처를 구분하였다. 그리고 계층2에서 추출된 제스처 정보를 바탕으로 CRF 모델을 이용하여 누적된 다수 투표 기법을 적용해 제스처의 종류를 판단하였다.

한편, 카메라 영상 외에 가속도 센서를 이용해 제스처를 인식한 연구도 있다. Akl의 연구[6]에서는 제스처 인식을 하기 위하여 시계열 데이터 집합을 군집화 하기 위해 군집화 알고리즘인 AP(Affinity Propagation)를 이용하였다. 제스처의 모델을 학습하기 위해 분류 학습 알고리즘인 DTW(Dynamic Time Wrapping)를 사용하였다. Akl의 연구는 가속도 센서를 이용해 얻은 데이터를 유사도에 따라 군집화 한다. 그리고 각 군집(Cluster)마다 대표(Exemplar)를 뽑아 대표와 테스트 샘플간의 DTW를 비교하여 제스처를 분류하는 방법을 사용하였다.

3. 팔 제스처 인식 시스템

3.1 시스템 개요

본 논문이 제안하는 팔 제스처 인식 시스템은, 1차원 정수 형태의 시계열 데이터를 HMM 학습 알고리즘에 적용하여 제스처별 모델을 학습하고, 시계열 데이터와 제스처별 HMM 모델을 통해 제스처를 인식한다.



(그림 1) 제안하는 팔 제스처 인식 시스템 구성

(그림 1)은 팔 제스처 인식 시스템의 전체 구성을 나타낸다. 이 시스템은 각각 전처리 단계, 학습 단계, 인식 단계의 크게 3단계로 나눌 수 있다. 먼저 전처리 단계에서는 키넥트 센서를 통해 양 팔의 어깨, 팔꿈치, 손목의 위치 좌표 데이터를 얻는다. 그리고 좌표 데이터를 8차원의 각도 데이터로 변환하고, HMM 모델 학습의 계산량을 줄이기 위하여 k-평균 군집화를 적용한다. k-평균 군집화를 이용하기 위해 군집의 수 또는 k의 개수를 결정해야 한다. k의 개수를 결정하기 위해 주성분 분석(PCA: Principal Component Analysis)을 이용해 도움을 받고자 하였다. k-평균 군집화를 적용하면 좌표 데이터는 1차원 정수 시계열 데이터가 된다. 전처리 단계의 결과는 2종류의 시계열 데이터로 각각 학습 단계와 테스트 단계에서 입력으로 사용되는 시계열 데이터이다.

학습 단계에서는 1차원의 학습 시계열 데이터를 입력받아 제스처별 HMM 모델을 학습한다. 마지막으로 인식 단계에서는 테스트 시계열 데이터를 이용해 제스처별 HMM 모델의 로그-우도 확률을 계산한다. 그리고 가장 높은 확률 값을 가진 제스처 모델을 테스트 시계열 데이터의 제스처로 인식한다.

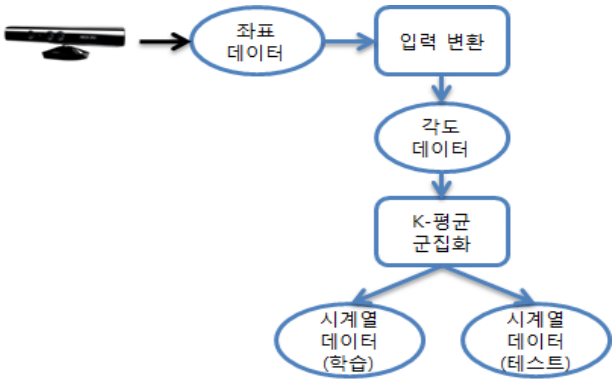
3.2 전처리 단계

(그림 2)은 전처리 단계의 흐름을 나타낸다. 키넥트 센서에서 측정된 팔의 어깨, 팔꿈치, 손목의 위치는 각각 x, y, z의 좌표로 표현된다. 이 좌표 데이터는 위치를 나타내는 공간적인 좌표로 이동이나 회전에 민감하게 반응한다. 그러므로 x, y, z로 이루어진 좌표 데이터를 ρ, θ, ϕ로 이루어진 각도 데이터로 변환한다. 각도 데이터로 변환하는 식은 식 (1)과 식(2)와 같다. 식 (1), (2)

$$\theta = \cos^{-1}\left(\frac{v_z}{\| v \|}\right) \qquad (1)$$

$$\phi = \tan^{-1}\left(\frac{v_y}{v_x}\right) \qquad (2)$$

에서 v는 어깨와 팔꿈치까지, 팔꿈치부터 팔목까지의 관절 간 벡

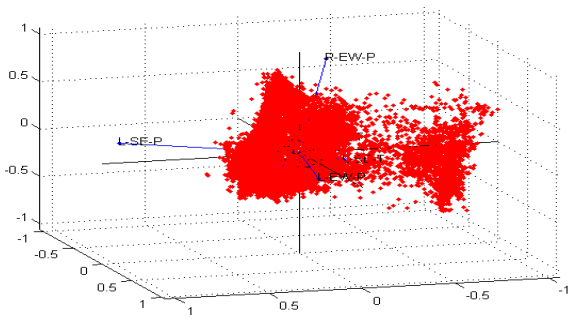


(그림 2) 전처리 단계 흐름도

터를 의미한다. θ는 v벡터의 극각(polar angle), ϕ는 v벡터의 방위각(azimuthal angle)이다. 식 (3)은 시간 t에서의 각도데이터이다. 예를 들어, θ_{RS_RE}은 오른팔 어깨와 오른팔 팔꿈치의 극각을 나타낸다.

$$x_t = (\theta_{RS\_RE},\ \phi_{RS\_RE},\ \theta_{RE\_RW},\ \phi_{RE\_RW},\ \theta_{LS\_LE},\ \phi_{LS\_LE},\ \theta_{LE\_LW},\ \phi_{LE\_LW})^T \qquad (3)$$

좌표 데이터를 각도 데이터로 변환한 다음 k-평균 군집화를 한다. 양 팔의 각도 데이터는 8차원의 실수 형태로 구성되어 있다. 다차원의 데이터를 사용하면 학습 알고리즘이 제스처별 모델을 학습하는데 필요한 계산량이 많아진다. 때문에 k-평균 군집화 과정을 통해 1차원의 정수 형태로 시계열 데이터를 바꾸어 준다. 각도데이터를 k-평균 군집화를 사용하여 군집화하기 위해서는 k 값을 설정해주어야 한다. 때문에 다차원 데이터를 3차원으로 축소하고, 데이터의 분포를 시각화 하여 k 값을 설정하는데 도움을 받고자 하였다. (그림 3)은 주성분 분석을 이용해 다차원 데이터를 3차원으로 차원 축소하여 각도 데이터의 분포를 시각화한 그림이다. 원점과 x축을 기준으로 0.8 정도의 위치에 각도 데이터가 몰려있는 것을 볼 수 있다.

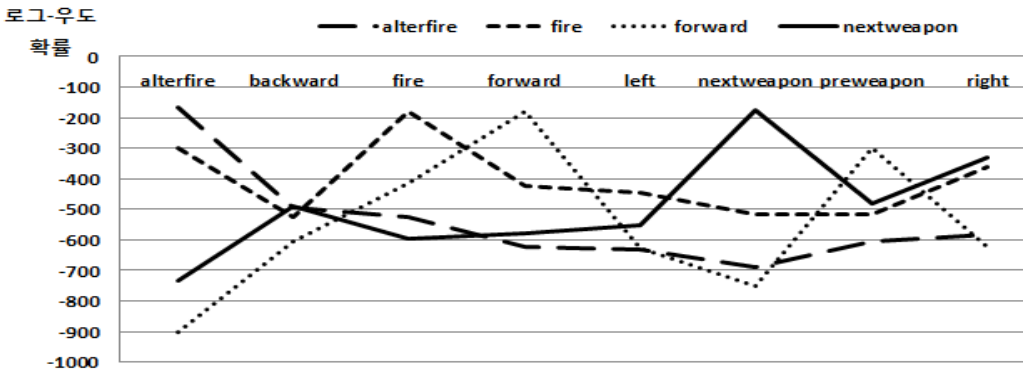


(그림 3) 주성분 분석을 이용한 각도 데이터 분포도

차원을 축소하여 각도 데이터를 시각화한 결과 군집이 뚜렷하게 드러나지 않아 k 값을 10, 20, 30으로 설정하였다. 그리고 8차원 각도 데이터에 k-평균 군집화를 적용하여 1차원 시계열 데이터를 추출하였다. k-평균 군집화 과정이 끝나면 학습 시계열 데이터와 테스트 시계열 데이터로 나뉜다. 학습 시계열 데이터는 제스처별 HMM 모델 학습에 사용되고, 테스트 시계열 데이터는 제스처 인식에 사용된다.

3.3 모델 학습 단계

특징 데이터를 추출 후 학습 시계열 데이터를 입력으로 받아 제스처별로 HMM 모델을 학습한다. HMM을 기반으로 한 Baum-Welch 알고리즘을 이용하여 학습 시계열 데이터가 발생할



(그림 5) 제스처 모델에 따른 로그-우도 확률

확률을 극대화 하는 모델을 학습하였다. 제스처의 개수만큼 모델을 학습한다. 각 제스처의 복잡한 정도에 따라 모델의 상태수를 최소 3개에서 최대 5개로 결정하였고, 초기 상태 값 및 상태 전이 확률은 임의의 값으로 설정하였다.

3.4 제스처 인식 단계

제스처 인식은 학습된 제스처별 HMM λ_{g^i} ($g^i \in G$, G는 HMM 모델의 집합)모델 과 테스트 시계열 데이터 집합 $X_f = [x_0, x_1, \ldots, x_f]$, $(0 \le f \le 100)$을 가지고 식 (4)를 이용한다. 학습된 모든 제스처별 HMM 모델 집합 중 로그-우도 확률이 가장 높은 모델을 구하고, 해당 인덱스의 제스처를 해당 제스처로 분류한다. 식 (4)는 같이 학습된 제스처별 HMM 모델이 있을 때, 입력된 테스트 시계열 데이터가 나타날 확률이 가장 큰 HMM 모델이 인식된 제스처이다.

$$\max(P(X|\ \lambda_{g^i})),\ \text{for all}\ g^i \in G \qquad (4)$$

4. 구현 및 실험

앞서 제안한 방법에 따라 Window 7 상에 팔 제스처 인식 시스템을 구현하였고, 이것을 이용해 팔 제스처 인식을 실험하였다. 먼저 실험에 필요한 데이터는 직접 수집한 데이터(Kyonggi 데이터 집합)와 Cornell 데이터 집합[7]이 있다. Kyonggi 데이터 집합을 수집하기 위해 키넥트 센서는 컴퓨터 모니터 상단에 설치하였고, Eclipse에 Kinect SDK 1.5를 설치하였다. 그리고 3명의 사람이 키넥트 센서로부터 1.5~2m 사이의 다양한 거리에서 촬영하였다.

(그림 4) Kyonggi 데이터 집합의 제스처

(그림 4)는 향후 슈팅 게임 언 리얼 토너먼트(Unreal Tournament)에 적용하기 위해 선정한 8가지 팔 제스처의 종류이다. 팔 좌표 데이터는 인당 10개씩 촬영하여 총 240개의 데이터이며, 초당 20프레임씩 5초간 촬영하였다. Cornell 데이터 집합은 군사 및 비행 제어를 위해 이용되는 팔 제스처이다. 15가지의 정적 제스처 데이터와 15가지의 동적 제스처 데이터로 구성되어 있으며, 본 논문에서는 15개의 동적 제스처 데이터를 이용하였다. 한 제스처 당 30개의 데이터씩 총 900개의 데이터로 구성되어 있다. 본 논문에서 제안하는 팔 제스처 인식 시스템은 위의 데이터를 가공하는 전처리 단계와 학습 단계 그리고 인식 단계로 구성되어 있다.

전처리 단계에서는 두 종류의 좌표 데이터 집합을 각도 데이터 집합으로 변경시킨 후, 이를 k값과 함께 k-평균 군집화 알고리즘의 입력 값으로 준다. k-평균 군집화의 경우 Java언어를 사용하여 구현하였다. 설정한 k값에 따라 각도 데이터 집합은 군집화를 거쳐 1부터 k의 범위를 가진 1차원 정수 형태의 시계열 데이터로 바뀐다. 시계열 데이터는 각각 20개의 학습 시계열 데이터와 10개의 테스트 시계열 데이터로 나뉜다.

학습 단계에서는 학습 시계열 데이터와 HMM 모델을 통해 제스처별 모델 학습을 수행한다. 학습에 사용되는 HMM 알고리즘은 Matlab(Matrix Laboratory) 프로그램을 이용하여 개발하였다. 학습 모델의 인덱스는 0부터 알파벳순으로 설정하였으며, 제스처별 HMM의 초기 모델은 난수로 생성하였다. 그리고 제스처별 HMM 모델의 상태 수는 제스처에 따라 3~5개로 결정하였다. 학습 단계의 결과로 Kyonggi 데이터 집합은 8개의 제스처이므로 8개의 HMM 모델이 생성되고, Cornell 데이터 집합은 15개의 제스처이므로 15개의 HMM모델이 생성된다.

인식 단계는 테스트 시계열 데이터와 학습된 모델을 이용하여 테스트 시계열 데이터의 제스처 종류를 인식하는 단계이다. 인식 단계의 입력 값은 테스트 시계열 데이터와 제스처별 HMM 모델의 사전확률, 상태 전이 확률, 관측 확률이며, 출력은 제스처 모델의 인덱스 값이다. 입력 값을 이용해 제스처 모델별 로그-우도 확률을 계산하며, 제스처별로 가장 큰 로그-우도 확률을 가진 모델을 테스트 시계열 데이터의 제스처로 인식한다. (그림 5)는 학습된 제스처 모델에 따른 시계열 데이터의 로그-우도 확률 값의 변화를 나타낸다. 학습된 8개의 HMM 모델에 테스트 시계열 데이터가 입력되었을 때 alterfire 시계열 데이터의 경우 alterfire 모델에서 가장 높은 로그-우도 확률을 가졌고 타 제스처 모델에서는 -500 이하의 낮은 확률 값을 가졌다. fire, forward, nextweapon 시계열 데이터 역시 해당하는 제스처 모델의 로그-우도 확률이 가장 높음을 보이고 있다.

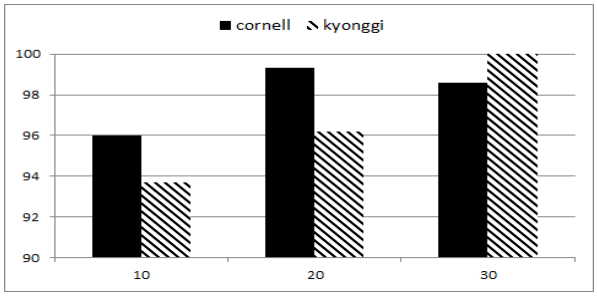
<표 1>은 k-평균 군집화 알고리즘의 군집 개수에 따른 Kyonggi 데이터 집합의 제스처별 성능을 나타낸다. 표를 보면 k의 개수가 증가할수록 성능이 점점 좋아지는 것을 볼 수 있다. 표에서 k의 개수가 10일 때와 20일 때, fire제스처와 left제스처의 성능이 다른 제스처에 비해서 낮게 나온다. 이는 fire는 alterfire

로, left는 preweapon으로 오분류를 했기 때문이다. 이는 서로 비슷한 방향과 모양이 원인으로 보인다.

<표 1> 군집 개수에 따른 HMM의 제스처별 인식률 변화

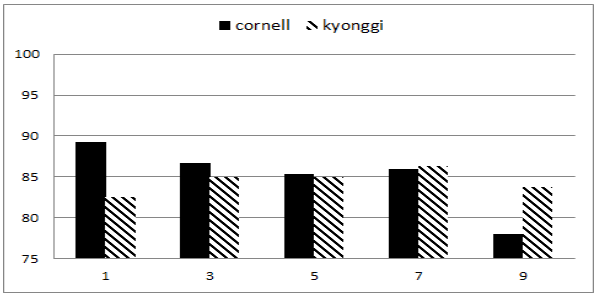
군집 개수 / 제스처	10	20	30
alterfire	9	10	10
backward	10	10	10
fire	9	8	10
forward	10	10	10
left	8	9	10
nextweapon	10	10	10
preweapon	10	10	10
right	9	10	10
확률	93.7%	96.2%	100%

(그림 6)의 그래프는 k의 개수에 따라, 테스트 시계열 데이터와 제스처별 HMM 모델을 이용해 제스처 인식을 수행한 결과이다. Cornell 데이터 집합과 Kyonggi 데이터 집합이 k-평균 군집화의 k의 개수에 따라 인식률이 다름을 알 수 있다. Kyonggi 데이터는 k의 개수가 30개일 때 , Cornell 데이터는 k의 개수가 30개일 때보다 20개일 때 성능이 가장 좋은 것으로 나타난다.



(그림 6) 군집 수에 따른 HMM의 평균 인식률 변화

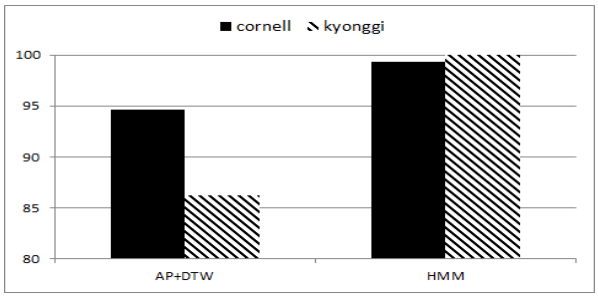
한편, 본 논문에서는 또 다른 시계열 데이터 학습방법인 DTW를 이용한 제스처 인식 성능을 분석해보기 위한 실험을 전개하였다. DTW는 두 시계열 데이터 간의 차이(difference) 혹은 거리(distance)를 계산해주는 효율적인 알고리즘이다. DTW를 이용한 가장 기본적인 제스처 인식방법은 모델을 학습하는 별도의 학습 단계없이 바로 인식 단계에서 DTW를 적용해 테스트용 시계열 데이터와 각각의 훈련용 데이터간의 유사도를 계산하고, 테스트 데이터에 가장 유사한 훈련용 데이터의 제스처 유형에 따라 테스트 데이터의 제스처를 결정하는 방식이다. 하지만, 훈련용 데이터 집합이 큰 경우 테스트 데이터와의 유사도 계산양이 증가하므로, 많은 기존 연구들에서는 실시간 인식 시간을 단축하기 위해 학습 단계에 AP(Affinity Propagation) 군집화를 적용하여 각 제스처별로 소수의 대표 데이터(exemplar)들을 선정한 다음, 인식 단계에서는 이 대표들과의 DTW 비교에 의해서만 테스트 데이터의 제스처를 결정하는 방법(AP+DTW)을 많이 이용한다. (그림 7)은



(그림 7) 대표 수에 따른 AP+DTW의 평균 인식률 변화

AP 군집화에 의해 선정되는 대표의 수에 따른 AP+DTW 제스처 인식 방법의 성능 실험 결과를 보여준다. Kyonggi 데이터 집합은 대표가 7일 때 가장 인식률이 좋다. 반면 Cornell 데이터 집합은 대표의 수가 증가할수록 인식률이 감소한다.

(그림 8)은 두 종류의 시계열 데이터 집합을 AP 군집화와 DTW로 제스처를 인식했을 때의 평균 인식률과 HMM을 이용해 제스처를 인식하였을 때의 평균 인식률을 나타낸다. AP+DTW의 경우 Cornell 데이터와 Kyonggi 데이터가 각각 95, 86%의 인식률을 보였다. HMM은 두 종류의 데이터 모두 99% 이상의 인식률을 보였다. 이는 비교 알고리즘에 비해 인식률이 4~15%정도 향상된 결과이다.



(그림 8) AP+DTW와 HMM의 평균 인식률

5. 결론

본 논문에서는 키넥트 센서를 이용하여 획득한 데이터를 통해 팔 제스처를 인식하였다. 이를 위해 k-평균 군집화 및 HMM 알고리즘을 이용하였다. k-평균 군집화 과정에서 성능 향상을 위해 k의 개수를 각각 10, 20, 30씩 늘려가며 실험을 했다. 그리고 이를 이용한 HMM과 AP 군집화와 DTW 알고리즘을 사용했을 때 제스처 인식 성능을 비교하였다. 결과적으로 AP 군집화와 DTW를 사용한 인식률 보다 k-평균 군집화와 HMM을 사용한 인식률이 더 높다는 결과가 나왔다. 향후 연속된 비디오 데이터를 이용해 제스처와 비 제스처를 구분하는 기술을 적용할 계획이다. 더불어 실제 슈팅게임에 제안한 모델을 적용하는 방향으로 연구를 진행할 계획이다.

참고문헌

[1] J. Sung, C. Ponce, B. Selman and A. Saxena, “Human Activity Detection from RGBD Images”, Proc. of Workshops at the 25th AAAI Conference on Artificial Intelligence Play, Activity, and Intent Recognition, 2011

[2] https://code.google.com/a/eclipselabs.org/p/jnect/

[3] Y. Li, “Multi-Scenario Gesture Recognition Using Kinect”, Proc. of the 17th International Conference on Computer Games, pp. 126-130, 2012

[4] H.K. Lee and J.H. Kim, “An HMM-Based Threshold Model Approach for Gesture Recognition”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 21, no. 10, pp.961-973, 1999

[5] S. Cho, H. Byun, H. Lee, J. Cha, “Arm Gesture Recognition for Shooting Games based on Kinect Sensor”, Jornal of KISSE : Software and Applications, vol. 39, no. 10, pp. 796~805, 2012

[6] A. Akl, et al. “A Novel Accelermometer-Based Gesture Recognition System”, IEEE Transaction on Signal Processing, vol. 59, no. 12, pp. 6197-6205, 2011

[7] http://pr.cs.cornell.edu/humanactivities/