

# Design of an Arm Gesture Recognition System Using Feature Transformation and Hidden Markov Models

Se-Kyeong Heo<sup>†</sup> · Ye-Seul Shin<sup>\*\*</sup> · Hye-Suk Kim<sup>\*\*\*</sup> · In-Cheol Kim<sup>\*\*\*\*</sup>

## ABSTRACT

This paper presents the design of an arm gesture recognition system using Kinect sensor. A variety of methods have been proposed for gesture recognition, ranging from the use of Dynamic Time Warping(DTW) to Hidden Markov Models(HMM). Our system learns a unique HMM corresponding to each arm gesture from a set of sequential skeleton data. Whenever the same gesture is performed, the trajectory of each joint captured by Kinect sensor may much differ from the previous, depending on the length and/or the orientation of the subject's arm. In order to obtain the robust performance independent of these conditions, the proposed system executes the feature transformation, in which the feature vectors of joint positions are transformed into those of angles between joints. To improve the computational efficiency for learning and using HMMs, our system also performs the k-means clustering to get one-dimensional integer sequences as inputs for discrete HMMs from high-dimensional real-number observation vectors. The dimension reduction and discretization can help our system use HMMs efficiently to recognize gestures in real-time environments. Finally, we demonstrate the recognition performance of our system through some experiments using two different datasets.

**Keywords :** Gesture Recognition, Kinect, Hidden Markov Models, k-Means Clustering

## 특징 변환과 은닉 마코프 모델을 이용한 팔 제스처 인식 시스템의 설계

허 세 경<sup>†</sup> · 신 예 슬<sup>\*\*</sup> · 김 혜 숙<sup>\*\*\*</sup> · 김 인 철<sup>\*\*\*\*</sup>

### 요 약

본 논문에서는 Kinect 센서를 이용한 팔 제스처 인식 시스템의 설계에 대해 소개한다. 제스처 인식을 위한 기존의 연구들에서는 동적 시간 왜곡(DTW)에서 은닉 마코프 모델(HMM)에 이르기까지 다양한 방법들이 적용되어 왔다. 본 논문에서 제안하는 제스처 인식 시스템은 Kinect 센서를 통해 얻을 수 있는 순차적인 팔 관절 위치 데이터로부터 각 제스처 별 고유한 은닉 마코프 모델을 학습한다. 동일한 제스처를 수행하더라도 Kinect 센서에 포착되는 각 관절의 위치 좌표 값들은 팔의 길이와 방향에 따라 크게 달라질 수 있다는 문제점이 있다. 본 논문에서 제안하는 시스템에서는 다양한 환경 조건에서도 높은 제스처 인식 성능을 얻기 위해, 팔 관절들의 좌표 값으로 구성된 특징 벡터를 팔 관절들 간의 각도 값으로 변환하는 특징 변환 과정을 수행한다. 또한, 본 시스템에서는 은닉 마코프 모델의 학습과 적용의 효율성을 높이기 위해, 고차원 실수 관측 벡터들에 k-평균 군집화를 적용하여 이산 은닉 마코프 모델들을 위한 1차원 정수 시퀀스들을 구한다. 이와 같은 차원 축소와 이산화를 통해, 실시간 환경에서도 은닉 마코프 모델들을 효율적으로 제스처 인식에 이용할 수 있다. 끝으로, 서로 다른 두 가지 데이터 집합을 이용한 실험을 통해, 본 논문에서 제안한 시스템의 높은 인식 성능을 입증해 보인다.

**키워드 :** 제스처 인식, Kinect, 은닉 마코프 모델, k-평균 군집화

## 1. 서 론

사람이 보다 편리한 방식으로 기계와 상호작용하기 위해 현재 연구되고 있는 기술들 중에서 특히 카메라 영상을 이용한 사람의 제스처 인식 기술이 큰 관심을 모으고 있다. 카메라 영상을 이용한 제스처 인식 기술은 원격 로봇 제어 [1], 컴퓨터 게임[2], 청각 장애인들을 위한 수화 인식[3] 등 매우 다양한 응용분야들에 활용될 수 있다. 특히 2010년 말에 등장한 Microsoft사의 Kinect 센서는 전통적인 RGB 영상 외에 깊이 영상(depth image)도 함께 제공해줄 수 있는

※ 본 연구는 경기도의 경기도지역협력연구센터사업의 일환으로 수행하였음.

※ 이 논문은 제39회 춘계학술발표대회에서 "키넥트 센서를 이용한 팔 제스처 인식 시스템의 설계"의 제목으로 발표된 논문을 확장한 것임.

† 준 회 원: 경기대학교 컴퓨터과학과 학사과정

\*\* 비 회 원: 경기대학교 컴퓨터과학과 학사과정

\*\*\* 준 회 원: 경기대학교 컴퓨터과학과 석사과정

\*\*\*\* 중 심 회 원: 경기대학교 컴퓨터과학과 교수

논문접수: 2013년 7월 4일

심사완료: 2013년 7월 29일

\* Corresponding Author: In-Cheol Kim(kic@kyonggi.ac.kr)

기능을 포함하고 있어, 최근 들어 Kinect 센서 및 깊이 영상을 이용한 제스처 인식 연구가 더욱 활기를 띠고 있다. 제스처 인식을 위한 기존의 많은 연구들에서는 은닉 마코프 모델(Hidden Markov Model, HMM)이나 조건부 랜덤 필드(Conditional Random Field, CRF)와 같은 확률 그래프 모델들을 이용하여 각 제스처별 고유 패턴을 학습하고 이들을 제스처 인식에 이용해왔다. 하지만 이러한 확률 그래프 모델들은 카메라 영상으로부터 실시간으로 추출되는 고차원 실수 벡터 형태의 시계열 데이터에 그대로 적용할 경우, 모델 학습과 적용에 매우 많은 계산 자원이 요구된다는 약점이 있다.

본 논문에서는 Kinect 센서를 이용한 팔 제스처 인식 시스템의 설계에 대해 소개한다. 본 논문에서 제안하는 제스처 인식 시스템은 Kinect 센서를 통해 얻을 수 있는 순차적인 팔 관절 위치 데이터로부터 각 제스처 별로 고유한 은닉 마코프 모델(HMM)을 학습한다. 본 논문에서 제안하는 시스템에서는 다양한 환경 조건에서도 높은 제스처 인식 성능을 얻기 위해, 팔 관절들의 위치 좌표 값으로 구성된 특징 벡터를 팔 관절들 간의 각도 값으로 변환하는 특징 변환 과정을 수행한다. 또한, 본 시스템에서는 은닉 마코프 모델의 학습과 적용의 효율성을 높이기 위해, 고차원 실수 관측 벡터들에 k-평균 군집화(k-means clustering)를 적용하여 이산 은닉 마코프 모델들을 위한 1차원 정수 시퀀스들을 구한다. 이와 같은 차원 축소(dimension reduction)와 이산화(discretization)를 통해, 실시간 환경에서도 은닉 마코프 모델들을 효율적으로 제스처 인식에 이용할 수 있다. 본 논문에서 제안한 시스템의 인식 성능을 분석하기 위해, 서로 다른 두 가지 데이터 집합을 이용한 실험을 수행하고, 그 결과를 소개한다.

2. 관련 연구

Lee의 연구[4]에서는 카메라 영상으로부터 사용자의 손 제스처를 인식해, 파워포인트(PowerPoint) 슬라이드 쇼를 제어하는 연구를 진행하였다. 이 연구에서는 연속적인 카메라 입력 영상으로부터 사용자의 제스처 동작과 비 제스처 동작을 구분하고자, 은닉 마코프 모델(HMM)을 이용한 적응적 임계치 모델(adaptive threshold model)을 제안하였다. Cho의 연구[5]에서는 팔 제스처 인식을 위해 2 계층 제스처 인식 모델을 제안하였다. 계층1에서는 은닉 마코프 모델을 이용하여 프레임별로 제스처와 비 제스처를 구분하는 소위 비 제스처 여과 처리를 수행하였다. 그리고 계층2에서는 계층1을 통해 제스처로 구분된 입력 데이터에 대해서 조건부 랜덤 필드(CRF) 모델을 이용하여 제스처의 종류를 판단하였다.

한편, Celebi의 연구[6]에서는 제스처 인식을 위해 시계열 분류 학습 알고리즘인 DTW(Dynamic Time Wrapping)를 사용하였다. DTW는 제스처 데이터와 같이 서로 다른 두 시계열 데이터 간의 유사도(similarity)를 효율적으로 계산해주는 알고리즘이다. DTW를 이용한 가장 기본적인 제스처

인식방법은 테스트 데이터를 학습용 제스처 데이터들과 일대일로 각각 비교하여 유사도가 가장 높은 학습용 제스처 데이터를 찾고, 해당 제스처로 테스트 데이터를 분류하는 것이다. 따라서 기본적인 DTW 기반의 제스처 인식 방법은 은닉 마코프 모델(HMM)이나 조건부 랜덤 필드(CRF) 모델과 같은 제스처별 모델을 학습하는 별도의 학습 단계가 필요하지 않다. Celebi의 연구에서는 제스처 분별력을 향상시키기 위해 특정 신체 관절들에 가중치를 부여한 Weighted DTW 알고리즘을 제스처 인식에 적용하였다. 반면에, Bautista[7]연구에서는 학습 단계에서 각 제스처별 훈련 데이터 집합에 대해 미리 GMM(Gaussian Mixture Model)을 구성해놓은 뒤, 인식 단계에서 이들을 이용한 Probabilistic DTW 알고리즘을 적용함으로써, 제스처 인식률을 높임과 동시에 인식 시간을 줄여 보려는 시도를 하였다.

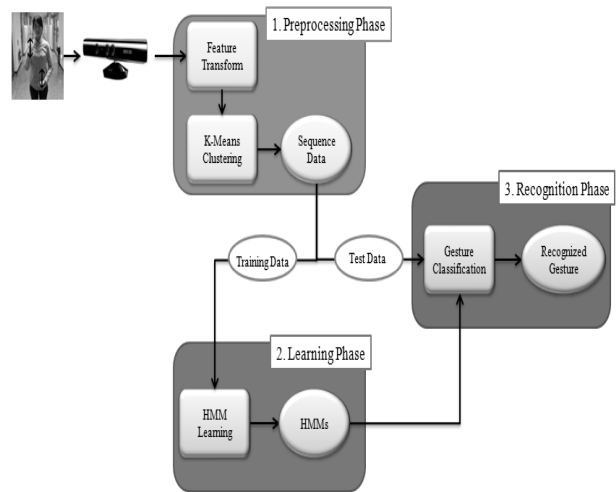


Fig. 1. Overall process of arm gesture recognition

3. 팔 제스처 인식 시스템

3.1 시스템 개요

본 논문에서 제안하는 팔 제스처 인식 시스템은 Kinect 센서 앞에서 사용자가 좌우 양 팔을 사용하여 미리 정해놓은 동적 제스처(dynamic gesture) 중 하나를 취할 때, Kinect 센서의 깊이 영상(depth image)들로부터 사용자의 양 팔 관절들의 위치 좌표를 추출하여 각 제스처를 나타내는 은닉 마코프 모델(Hidden Markov Model, HMM)을 학습하고 이를 이용함으로써 실시간으로 사용자가 취하는 제스처를 인식해내는 시스템이다.

팔 제스처 인식 시스템의 전체 처리 과정은 Fig. 1과 같이 전처리 단계(preprocessing phase), 학습 단계(learning phase), 인식 단계(recognition phase) 등 크게 3 단계로 구성된다. 먼저 전처리 단계에서는 초당 20 프레임(frame/sec)의 속도로 얻어지는 Kinect 센서의 깊이 영상으로부터 양 팔의 어깨, 팔꿈치, 손목 등 각 주요 관절들(joint)의 위치 좌표로 구성된 특징 벡터를 추출한다. 그리고 이렇게 추출

한 관절 좌표 특징 벡터를 팔 길이와 촬영 방향에 민감하지 않으면서 제스처를 잘 분별할 수 있도록 인접한 두 관절들 간의 각도 데이터로 특징 변환(feature transformation)을 수행한다. 이어서, 변환된 고차원 실수 특징 벡터들에 k-평균 군집화(k-means clustering)를 적용하여 유사한 특징 벡터들을 1~k 사이의 해당 군집 번호로 변환하는 이산화(discretization)를 수행한다. 제스처 동영상으로부터 5초 동안 약 100 개의 영상 프레임들을 얻어내고 이들 각 프레임에 대해 이와 같은 전처리 단계를 거침으로써, 5초 간격으로 길이가 100인 1차원 정수 시퀀스 형태의 제스처 입력 데이터를 생성한다.

학습 단계에서는 각 제스처를 나타내는 다수의 시계열 훈련 데이터 집합을 이용하여, 제스처별 은닉 마코프 모델(HMM)을 학습한다. 이때 학습에 사용되는 은닉 마코프 모델은 k개의 서로 다른 관측들 중 하나를 가질 수 있는 이산 은닉 마코프 모델이며, 학습 알고리즘으로는 EM (Expectation & Maximization) 알고리즘의 한 유형인 Baum-Welch 알고리즘을 이용한다. 마지막으로 인식 단계에서는 Kinect 센서를 통해 실시간으로 입력되는 테스트 시계열 데이터에 대해 각 제스처별 은닉 마코프 모델을 적용하여, 최대 로그-우도(maximum log likelihood)를 갖는 모델의 제스처로 판별한다.

3.2 전처리 단계

Fig. 2는 전처리 단계의 세부 과정을 나타내고 있다. Microsoft사의 API를 이용하여 Kinect 센서의 깊이 영상으로부터 제스처를 수행하는 사람의 양 팔 어깨, 팔꿈치, 손목 등 각 주요 팔 관절의 3차원 위치 데이터를 x, y, z의 실수 좌표 형태로 얻을 수 있다. 이렇게 구해지는 팔 관절의 위치 좌표 데이터는 Kinect 센서의 관점에서 제스처 수행자 팔의 상대적인 위치와 방향, 그리고 팔의 길이 등 다양한 요인들에 매우 민감하게 변화하는 특성을 가지므로, 다양한 환경 조건에서도 제스처들을 정확하게 분별하기 위한 특징(feature)들로 부족하다.

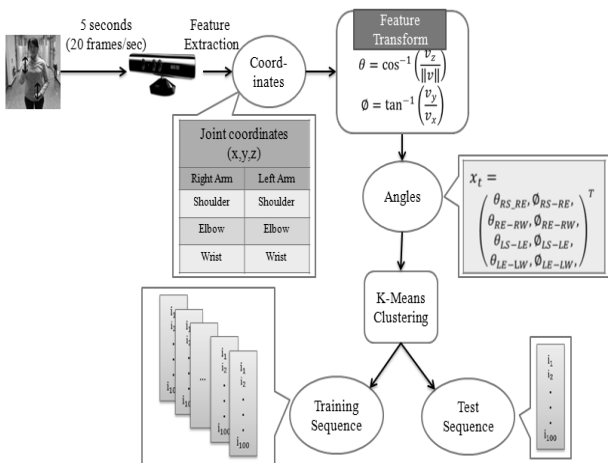


Fig. 2. Preprocessing phase

따라서 본 논문에서 제안하는 제스처 인식 시스템에서는 x, y, z로 이루어진 각 팔 관절의 위치 좌표 데이터를  $\rho, \theta, \phi$ 로 이루어진 관절들 간의 각도 데이터로 변환하는 특징 변환(feature transformation) 과정을 수행한다. 각도 데이터로 변환하는 식은 식 (1)과 식(2)와 같다. 식 (1), (2)

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

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

에서  $v$ 는 어깨와 팔꿈치까지, 팔꿈치부터 팔목까지의 관절 간 벡터를 의미한다.  $\theta$ 는  $v$ 벡터의 극각(polar angle),  $\phi$ 는  $v$  벡터의 방위각(azimuthal angle)이다. 식 (3)은 이러한 변환을 거쳐 구해지는 시간 t의 특징 벡터로서, 각 특징은 이웃한 두 관절들이 이루는 각도를 나타낸다. 예를 들어,  $\theta_{RS,RE}$  은 오른팔 어깨와 오른팔 팔꿈치의 극각을 나타낸다.

$$x_t = (\theta_{RS,RE}, \phi_{RS,RE}, \theta_{RE,RW}, \phi_{RE,RW}, \theta_{LS,LE}, \phi_{LS,LE}, \theta_{LE,LW}, \phi_{LE,LW})^T \tag{3}$$

지금까지 설명한 전처리 과정을 다시 정리하면, Kinect 센서를 통해 실시간 제스처 동영상으로부터 초당 20 개 프레임(frame)씩 깊이 영상을 입력하고, 각 프레임에 대해 팔 관절들의 위치 좌표들로 구성된 특징 벡터를 추출한 뒤, 이 특징 벡터에 다시 이웃한 관절들 간의 각도들로 변환하는 특징 변환 과정을 적용함으로써, 식 (3)과 같은 8차원 실수 특징 벡터를 생성한다. 하지만, 이러한 고차원 실수 벡터들의 시퀀스를 그대로 관측 열(observation sequence)로 삼아 제스처별 연속 은닉 마코프 모델들을 학습하고 이것을 다시 제스처 인식에 적용하는 방안은 계산 부담이 매우 커져 실시간 제스처 인식에 적합하지 않다. 따라서 본 논문에서 제안하는 시스템의 전처리 단계에서는 k-평균 군집화(k-means clustering)를 적용하여 유사한 8차원 실수 특징 벡터들을 묶어 k 개의 군집으로 나누고, 새로운 특징 벡터가 입력되면 이 벡터가 속한 군집의 번호(1~k 범위의 정수)로 이 고차원 실수 벡터를 대신 교체하는 차원 축소(dimension reduction) 및 이산화(discretization) 과정을 수행한다.

k-평균 군집화를 적용하기 위해서는 미리 생성하고자 하는 군집의 수(k)를 설정해주어야 하는데, 이것은 입력되는 고차원 특징 벡터들의 분포에 따라 적절히 결정되어야 한다. 일반적으로 입력 특징 벡터들이 중앙에 집중적으로 몰려서 분포할 때는 군집의 수(k)를 작게 설정해주어도 되고, 입력 특징 벡터들의 분포가 넓게 분산될수록 군집의 수(k)를 크게 설정해준다. 한편, 이 군집의 수(k)는 제스처 학습과 인식 단계에 적용될 이산 은닉 마코프 모델의 관측 정밀도와도 직접 연관된다. 즉, 군집의 수(k)가 증가할수록 이산

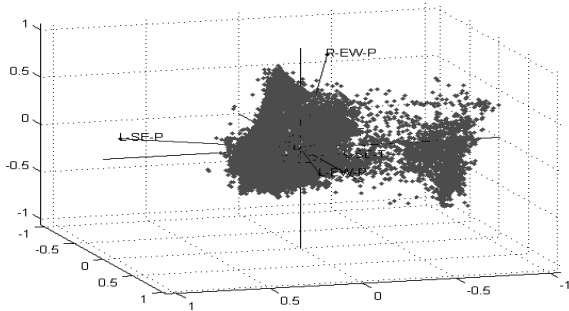


Fig. 3. PCA-based visualization of high-dimensional feature vectors

은닉 마코프 모델의 관측 정밀도도 비례해서 증가하므로 각 제스처를 세밀하게 모델링할 수 있다는 장점은 있으나, 대신 모델 학습과 인식의 계산 복잡도도 증가하는 단점이 있다. 반대로 군집의 수(k)가 감소할 때는 이와 반대되는 결과가 발생한다. 본 시스템에서는 특징 벡터들의 적절한 군집 수(k)를 결정하기 위해, 훈련 목적으로 수집된 8차원 실수 특징 벡터들을 Fig. 3과 같이 주성분 분석(Principal Component Analysis, PCA)을 통해 3차원 공간위의 점들로 투영하여 분포를 살펴보았다. 본 시스템에서는 이와 같은 분석을 통해 군집의 수(k)를 20~30으로 설정하였다.

전처리 단계의 마지막 과정에서는 앞서 설명한 방식대로 각 영상 프레임별로 특징 벡터와 이를 대신하는 군집 번호를 생성한 뒤, 이들을 5초 간격으로 약 100 개씩 순차적으로 모아 길이가 100인 1차원 정수 시퀀스 데이터를 생성한다. 이렇게 생성된 시퀀스 데이터는 각 제스처별 이산 은닉 마코프 모델들을 학습하기 위한 훈련 데이터와 제스처 인식을 위한 테스트 데이터로 각각 사용된다.

3.3 모델 학습 단계

모델 학습 단계에서는 제스처별로 수집된 다수의 훈련 데이터 집합을 이용하여, 각 제스처를 판별하는데 이용할 은닉 마코프 모델(HMM)들을 학습한다. 각 훈련 데이터는 전처리 단계를 통해 얻어지는 1차원 군집 색인 열(cluster index sequence)과 제스처 라벨(label)의 쌍으로 구성된다. 이때 군집 색인은 8차원 실수 특징 벡터가 속한 군집의 번호를 나타낸다. 각 제스처를 나타내는 확률 그래프 모델인 은닉 마코프 모델은 관측 변수(observation variable)와 은닉 상태 변수(hidden state variable)로 구성된다. 본 시스템에서 사용하는 모델의 은닉 상태 변수는 서로 다른 3~5개의 상태(state)들 중 하나를 값으로 가질 수 있고, 관측 변수는 서로 다른 20~30개의 관측(observation)들 중 하나를 값으로 가질 수 있는 이산 은닉 마코프 모델이며, 연결 형태는 Fig. 4와 같이 상태 전이가 왼쪽에서 오른쪽으로 한쪽 방향으로만 허용되는 좌-우 모델(left-right model)이다. 하나의 은닉 마코프 모델을 구성하는 3가지 파라미터(parameter)들은 초기 상태 확률 분포, 상태 전이 확률 분포, 관측 확률 분포들이며, 학습 단계에서는 훈련 데이터집합을 이용하여 각 제

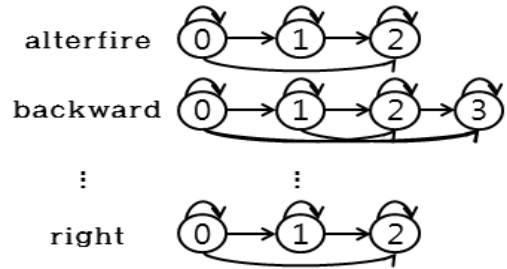


Fig. 4. Hidden Markov models for gesture recognition

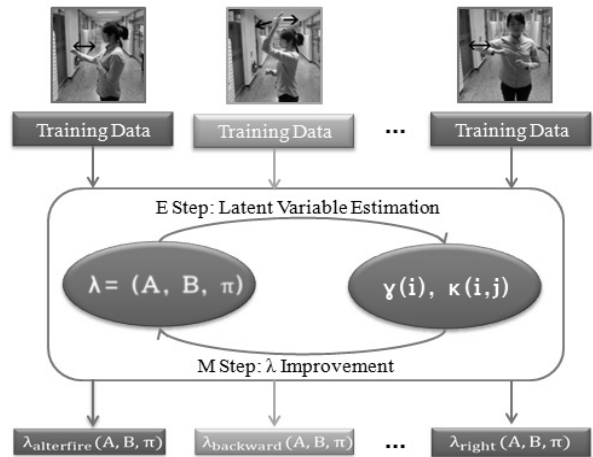


Fig. 5. Baum-Welch algorithm for learning HMMs

스처별 은닉 마코프 모델의 이 3가지 파라미터들의 최적치를 학습하는 것이 목적이다.

각 제스처별 은닉 마코프 모델의 학습을 위해서는 Fig. 5에 표현한 Baum-Welch 알고리즘을 이용한다.

Baum-Welch 알고리즘은 EM(Expectation & Maximization) 학습 알고리즘의 일종으로서, Fig. 5와 같이 은닉 변수들인  $\gamma(i)$ 와  $\kappa(i,j)$ 를 추정하는 E 단계와 모델의 파라미터들인  $\lambda=(A, B, \pi)$ 를 개선시키는 M 단계를 계속 반복함으로써 은닉 마코프 모델을 위한 최적의 파라미터들을 학습하는 알고리즘이다. 본 연구에서는 Baum-Welch 알고리즘을 이용하여 각 제스처별 은닉 마코프 모델을 학습하기에 앞서, 제스처의 복잡도에 따라 해당 모델의 상태수를 3~5개로 차등적으로 설정해주었고, 각 모델의 초기 상태 확률 분포, 상태 전이 확률 분포, 관측 확률 분포 등 3가지 파라미터들의 초기치는 난수 발생기를 이용하여 임의로 설정해주었다.

3.4 제스처 인식 단계

제스처 인식 단계에서는 전처리 단계를 거쳐 새로운 테스트용 시계열 데이터가 입력되면 여기에 각 제스처별 은닉 마코프 모델들을 적용하여, 각 모델별로 해당 테스트 데이터의 발생 가능성을 나타내는 로그 우도(log likelihood)를 계산한다. 그리고 최대 로그 우도를 가지는 모델을 찾아 그 모델이 나타내는 제스처로 테스트 데이터를 판별한다. 식 (4)는 각 제스처  $g^i \in G$ 별로 학습된 은닉 마코프 모델  $\lambda_{g^i}$ 들

을 이용하여 테스트 시계열 데이터  $X$ 에 대한 로그 우도  $P(X|\lambda_{g^i})$ 를 계산한 뒤, 이때 최대 로그 우도를 갖는 모델  $\lambda_{g^i}$ 을 찾아 해당 제스처  $g^*$ 로 인식하는 방법을 수식화한 것이다.

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

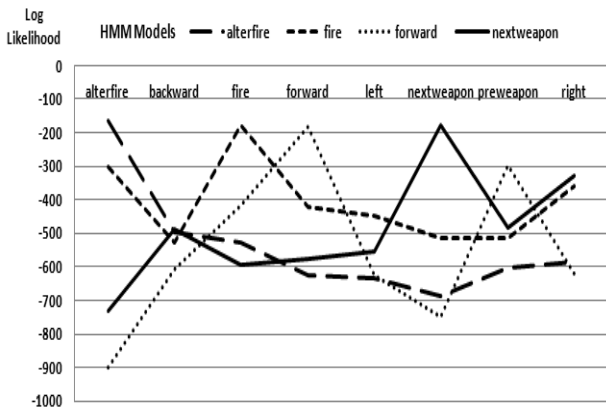


Fig. 6. Gesture recognition using learnt HMMs

Fig. 6은 한 실험자가 그림의 왼쪽에서 오른쪽으로 진행하면서 서로 다른 8 가지 제스처들을 차례대로 취하는 동안 각 제스처별로 학습된 은닉 마코프 모델들이 계산한 로그 우도의 변화를 나타내고 있다. 예컨대, 그림의 맨 왼쪽에는 실험자가 alterfire 제스처를 수행했을 때, 인식 시스템 내부의 각 은닉 마코프 모델들의 로그 우도를 비교해 보여주고 있다. Fig. 6이 나타내는 예의 경우에는 최대 로그 우도를 갖는 제스처, 즉 인식 시스템이 내부적으로 추정된 제스처도 실험자가 수행한 실제 제스처와 일치하는 alterfire임을 그림에서 확인할 수 있다.

#### 4. 구현 및 실험

앞서 제안한 방법에 따라 Window 7 상에 팔 제스처 인식 시스템을 구현하였고, 이것을 이용해 팔 제스처 인식을 실험하였다. 실험에 사용한 제스처 데이터 집합은 본 연구를 위해 자체적으로 수집한 Kyonggi 데이터 집합과 연구용으로 공개되어 있는 Cornell 데이터 집합[8]이다.

Kyonggi 데이터 집합은 Fig. 7과 같은 Unreal Tournament[9] 슈팅 게임을 수행하기 위해 Kinect를 통해 수집된 총 8 가지 동적 팔 제스처 데이터들을 포함하고 있다. Kyonggi 데이터 집합에 포함된 동적 팔 제스처 종류는 Fig. 8과 같다. Kyonggi 데이터 집합은 3명의 서로 다른 실험자들이 Kinect 센서로부터 1.5~2m 사이의 다양한 거리에서 8 가지 제스처 각각을 10회씩 수행하면서 얻은 총 240 개의 제스처 데이터들을 포함하고 있다. 각 제스처 데이터는 해당 제스처를 수행하는 동안 Kinect로 초당 20프레임씩



Fig. 7. Screenshot of Unreal Tournament game



Fig. 8. Gestures included in Kyonggi dataset



Fig. 9. Gestures included in Cornell dataset

5초간 촬영하면서 Jnect API[10]를 통해 얻은 양 팔의 총 6 개 관절(오른/왼쪽 어깨, 오른/왼쪽 팔꿈치, 오른/왼쪽 손목)의 3차원 위치 좌표(x, y, z)의 변화 시퀀스이다.

Cornell 데이터 집합은 주로 군사 목적의 비행 제어를 위한 팔 제스처 데이터들을 포함하고 있다. 이 데이터 집합에는 15가지 종류의 정적 팔 제스처와 15가지 종류의 동적 팔 제스처를 나타내는 총 900개의 제스처 데이터들이 포함되어 있다. 각 제스처 데이터는 Kyonggi 데이터 집합과 마찬가지로, 해당 제스처를 수행하는 동안 변화한 양 팔 주요 관절들의 3차원 위치 좌표들이다. 본 논문에서는 Kyonggi 데이터 집합과의 비교를 위해 이들 중 동적 팔 제스처들을 나타내는 450개의 데이터들만을 실험에 이용하였다.

본 논문에서 제안하는 팔 제스처 인식 시스템은 앞서 설명한 바와 같이 크게 전처리 단계와 모델 학습 단계, 그리고 제스처 인식 단계로 구성되며, 부분적으로 Java와 Matlab을 혼용하여 구현하였다. 전처리 단계에서는 먼저 좌

표 데이터를 각도 데이터로 변환한 뒤, 다시 k-평균 군집화를 적용하여 1차원 정수 형태의 시계열 데이터로 변환하였다. 전처리가 완료된 시계열 데이터 집합은 제스처별로 각각 20개의 학습용 데이터와 10개의 테스트용 데이터로 나누어 실험에 이용하였다.

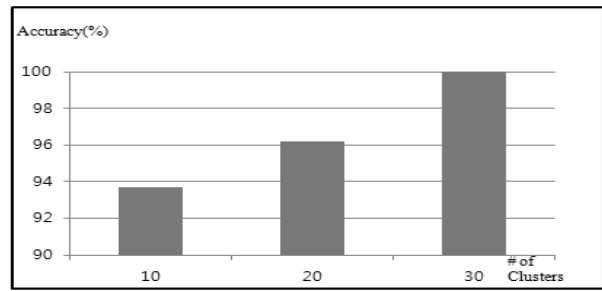
모델 학습 단계에서는 학습용 데이터로부터 각 제스처별로 고유한 은닉 마코프 모델(HMM)을 하나씩 생성하였다. 은닉 마코프 모델 학습에 이용한 학습 알고리즘은 EM(Expectation-Maximization) 알고리즘의 한 종류인 Baum-Welch 알고리즘이다. 제스처별 은닉 마코프 모델의 상태 수는 제스처의 복잡도에 따라 3~5개로 설정하였으며, 각 모델의 초기 파라미터 값들은 난수 생성기를 이용하여 임의로 설정하였다. 모델 학습의 결과로 Kyonggi 데이터 집합의 경우는 총 8개의 은닉 마코프 모델들이 생성되었고, Cornell 데이터 집합의 경우는 총 15개의 은닉 마코프 모델들이 생성되었다.

제스처 인식 단계에서는 학습된 은닉 마코프 모델들을 이용하여, 새로 입력되는 테스트 데이터의 제스처 종류를 인식하도록 구현하였다. 즉, 제스처별 은닉 마코프 모델들과 테스트용 신규 데이터를 제스처 인식 단계의 입력으로 주었으며, 주어진 테스트 데이터에 대해 최대 로그-우도를 갖는 모델의 제스처 번호를 출력으로 생성하였다.

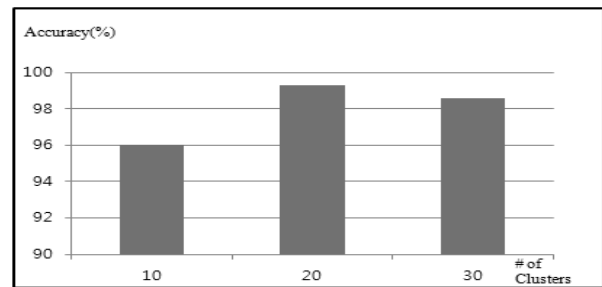
Table 1은 Kyonggi 데이터 집합에 대해 본 논문에서 제안한 제스처 인식 시스템의 인식 성능을 실험한 결과를 보여주고 있다. 차원 축소와 이산화를 위한 k-평균 군집화가 제스처 인식 성능에 미치는 영향도 함께 분석해보기 위해, 군집 개수 k를 10, 20, 30으로 증가시키면서 각 경우 대한 제스처별 인식 정확도를 측정하여 보았다. 그 결과 본 논문에서 제안한 시스템이 93.7% ~ 100%의 높은 인식 정확도를 보였다. 또한, Kyonggi 데이터 집합의 경우, 전체적으로 군집 개수 k가 증가할수록 인식 정확도도 함께 증가한 것을 확인할 수 있었다. 하지만, 실험 중에 fire 제스처를 alterfire로, left 제스처를 preweapon으로 오분류하는 경우가 가끔 발생하였는데, 이로 인해 fire제스처와 left제스처에 대한 인

Table 1. Experimental result for Kyonggi dataset

# of Clusters \ Gestures	10	20	30
alterfire	9/10	10/10	10/10
backward	10/10	10/10	10/10
fire	9/10	8/10	10/10
forward	10/10	10/10	10/10
left	8/10	9/10	10/10
nextweapon	10/10	10/10	10/10
preweapon	10/10	10/10	10/10
right	9/10	10/10	10/10
Accuracy(%)	93.7	96.2	100



(a) Kyonggi dataset



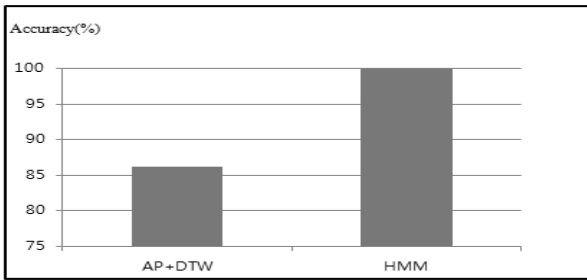
(b) Cornell dataset

Fig. 10. Average recognition accuracy

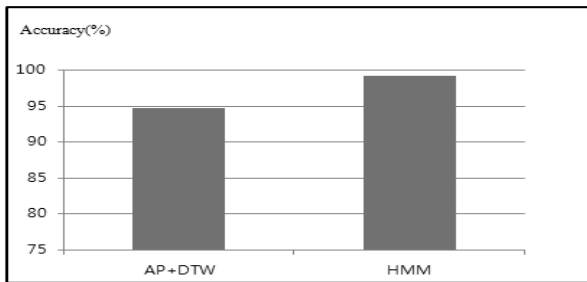
식률이 다른 제스처에 비해 약간 낮게 측정되었다. 추정되는 오분류의 원인으로는 fire와 alterfire, left와 preweapon은 모두 서로 비슷한 팔 모양과 진행 방향을 가지기 때문인 것으로 추정된다.

Fig. 10의 (a)와 (b)는 각각 Kyonggi 데이터 집합과 Cornell 데이터 집합에 대해, 본 논문에서 제안한 제스처 인식 시스템이 보여준 인식 성능을 나타낸다. 각각의 막대 그래프는 군집 개수 k를 10, 20, 30으로 증가했을 때 평균 제스처 인식 정확도의 변화를 나타내고 있다. Fig. 10의 (a)에서 보듯이 Kyonggi 데이터 집합의 경우는 군집 개수 k가 증가함에 따라 제스처 인식 정확도도 증가하였지만, Fig. 10의 (b)에서 보듯이 Cornell 데이터 집합의 경우는 군집 개수에 비례해서 인식 정확도가 증가하지 않고 군집 개수 k가 20일 때 98.2%의 최고 인식 정확도를 보였다.

한편, Fig. 11은 본 논문에서 제안한 은닉 마코프 모델(HMM) 기반의 제스처 인식 시스템의 성능을 또 다른 제스처 인식 방법(AP+DTW)과 비교해보기 위한 실험 결과를 보여준다. 기본적인 DTW 기반의 제스처 인식 방법은 제스처 인식 단계에서 대용량의 학습 데이터들과 테스트 데이터를 모두 비교하여 유사도를 계산해야 하므로 실시간 제스처 인식이 어렵다. 이러한 문제점을 극복하고자, AP+DTW를 이용한 제스처 인식 방법에서는 학습 단계에서 학습용 제스처 데이터들에 AP(Affinity Propagation) 군집화를 적용함으로써 미리 각 제스처별 소수의 대표(exemplar) 학습 데이터들을 선정해둔 뒤, 제스처 인식 단계에서는 테스트 데이터를 선정된 소수의 대표 데이터들과만 비교함으로써 제스처를 판별한다. 따라서 AP+DTW 기반의 제스처 인식 방법은



(a) Kyonggi dataset



(b) Cornell dataset

Fig. 11. Comparison: AP+DTW vs HMM

본 연구와 같이 실시간 제스처 인식에 많이 이용되는 방법이다. Fig. 11은 두 인식 시스템이 최고의 성능을 낼 수 있도록 각자의 매개변수(군집 개수, 대표 개수 등)들을 설정한 상태에서 실험한 결과를 보여준다. Fig. 11의 (a)와 (b)에서 보듯이 AP+DTW 기반의 제스처 인식 시스템은 Kyonggi 데이터 집합과 Cornell 데이터 집합에 대해, 86%와 95, 86%의 인식률을 보였다. 반면에, 본 논문에서 제안한 은닉 마코프 모델(HMM) 기반의 인식 시스템은 두 데이터 집합 모두에 대해, 99% 이상의 뛰어난 인식률을 보였다. 이것은 AP+DTW 기반의 인식 시스템에 비해 4~15%정도 더 높은 인식 성능을 보인 것으로서, 이를 통해 본 시스템의 상대적 성능 우수성을 확인할 수 있었다.

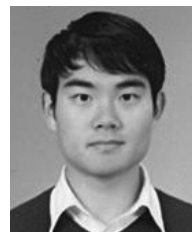
### 5. 결 론

본 논문에서는 Kinect 센서를 이용한 동적 팔 제스처 인식 시스템의 설계에 대해 소개하였다. 본 논문에서 제안하는 은닉 마코프 모델(HMM) 기반의 팔 제스처 인식 시스템에서는 다양한 환경 조건에서도 높은 제스처 인식 성능을 얻기 위해, 팔 관절들의 좌표 값으로 구성된 특징 벡터를 팔 관절들 간의 각도 값으로 변환하는 특징 변환 과정을 수행하였다. 또한, 본 시스템에서는 은닉 마코프 모델의 학습과 적용의 효율성을 높이기 위해, 고차원 실수 관측 벡터들에 k-평균 군집화를 적용하여 차원 축소와 이산화를 수행하였다. Kyonggi 데이터 집합과 Cornell 데이터 집합과 같은 서로 다른 두 가지 데이터 집합을 이용한 실험을 통해, 본 논문에서 제안한 시스템의 높은 인식 성능을 확인하였다.

향후 본 논문에서 소개한 팔 제스처 인식 시스템을 실제 Unreal Tournament 슈팅 게임과 연동하는 연구를 추가적으로 진행할 계획이다.

### 참 고 문 헌

- [1] M. Peris, K. Fukui, "Both-Hand Gesture Recognition Based on KOMSM with Volume Subspaces for Robot Teleoperation", Proceedings of IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems, 2012.
- [2] Y. Li, "Multi-Scenario Gesture Recognition Using Kinect", Proc. of the 17th International Conference on Computer Games, pp.126-130, 2012
- [3] S. Lang, M. Block, R. Rojas, "Sign Language Recognition Using Kinect", Artificial Intelligence and Soft Computing, Lecture Notes in Computer Science, Vol.7267, 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", Journal of KISSE : Software and Applications, Vol.39, No.10, pp.796-805, 2012.
- [6] S. Celebi, et al, "Gesture Recognition Using Skeleton Data with Weighted Dynamic Time Warping", Proceedings of International Conference on Computer Vision Theory and Applications, 2013.
- [7] M. A. Bautista, et al, "Probability-Based Time Warping for Gesture Recognition on RGB-D Data", Proceedings of International Conference on Pattern Recognition Workshops, 2012.
- [8] <http://pr.cs.cornell.edu/humanactivities/handgesture/handgesturedata.php>
- [9] <http://www.unrealtournament.com>
- [10] <http://code.google.com/a/eclipselabs.org/p/jnct>



#### 허 세 경

e-mail : tprudzzang@kyonggi.ac.kr

2008년~현 재 경기대학교 컴퓨터과학과  
학사과정

관심분야 : 인공지능, 기계학습, 제스처 인식



#### 신 예 슬

e-mail : ysulshin@kyonggi.ac.kr

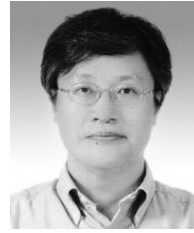
2010년~현 재 경기대학교 컴퓨터과학과  
학사과정

관심분야 : 인공지능, 기계학습, 제스처 인식



**김혜숙**

e-mail : chfls612@kyonggi.ac.kr  
2013년 경기대학교 컴퓨터과학과(학사)  
2013년~현재 경기대학교 컴퓨터과학과  
석사과정  
관심분야: 인공지능, 지능로봇, 컴퓨터비전



**김인철**

e-mail : kic@kyonggi.ac.kr  
1985년 서울대학교 수학과(학사)  
1987년 서울대학교 전산학과(이학석사)  
1995년 서울대학교 전산학과(이학박사)  
1996년~현재 경기대학교 컴퓨터과학과  
교수

관심분야: 인공지능, 기계학습, 지능시스템