# 연속적인 전신 제스처에서 강인한 행동 적출 및 인식

박아연, 신호근, 이성환
고려대학교 정보통신대학 인공시각연구센터 / 컴퓨터학과
{aypark, hkshin, swlee}@image.korea.ac.kr

# Robust Gesture Spotting and Recognition in Continuous Full Body Gesture

A.-Y. Park, H.-K. Shin and S.-W. Lee
Dept. of Computer Science and Engineering, Korea University

요 약

강인한 행동 인식을 하기 위해서는 연속적인 전신 제스처 입력에서부터 의미 있는 부분만을 분할하는 기술이 필요하다. 하지만 의미 없는 행동을 정의하고, 모델링 하기 어렵기 때문에, 연속적인 행동에서 중요한 행동만을 분할한다는 것은 어려운 문제이다. 본 논문에서는 연속적인 전신 행동의 입력으로부터 의미있는 부분을 분할하고, 동시에 인식하는 방법을 제안한다. 의미 없는 행동을 제거하고, 의미 있는 행동만을 적출하기 위해 garbage 모델을 제안한다. 이 garbage 모델에 의해 의미 있는 부분만 HMM의 입력으로 사용되어지며, 학습되어진 HMM 중에서 가장 높은 확률 값을 가지는 모델을 선택하여, 행동으로 인식한다. 제안된 방법은 20명의 3D motion capture data와 Principal Component Analysis를 이용하여 생성된 80개의 행동 데이터를 이용하여 평가하였으며, 의미 있는 행동과, 의미 없는 행동을 포함하는 연속적인 제스처 입력열에 대해 98.3%의 인식률과 94.8%의 적출률을 얻었다.

## 1. 서 론

제스처(gesture)는 일상생활에서 인간이 몸이나, 손, 얼굴들을 이용하여 의미 있는 의사를 표현하는 방법이며, 따라서 컴퓨터 비젼 기술을 이용하여, 자동적으로 인간의 행동을 인식하는 것은 기계와 인간사이의 새로운 인터페이스로 작동할 수 있으며 이러한 연구가 활발히 연구되어왔다[1]. 예측할 수 없는 입력신호로부터 이를 인식하는 기법을 "패턴 적출"이라 한다, 제스처 인식은 제스처의 시작점과 끝점을 찾기 위한 방법을 필요로 한다는 면에서 패턴 적출 응용중의 하나라고 볼 있다. 따라서 본 논문에서는 각각의 구성요소들의 특징을 그룹화하고 제스처의 시작과 끝점을 인식하기 위해 은닉 마르코프 모델 인식 방법을 제안하고 그 성능을 평가하고자 한다.

인간 전신의 행동을 분석하는 방법은 크게 형판 정합(template matching) 방법과 시간의 변화를 고려한 은닉 마르코프(Hidden Markov Model) 방법이 널리 사용되고 있다. 형판 정합의 대표적인 방법[2]은 인간 행동의 변화를 시간에 따른 영상(Motion History Image-MHI)으로 표현하고, 입력 MHI와 데이터베이스에 저장 되어진 MHI를 비교함으로써 행동을 인식하게 된다. 하지만 이러한 방법은 배경이 변화할 경우 MHI의 생성의 오류가 발생하며, 이러한 문제는 인식률에 영향을 미치게 된다. 또한 이 방법의 경우에는 행동의 시작과 끝은 판단하는데 어려움 있다는 단점을 가지고 있다. 반면 은닉 마르코프 모델의 경우는 시간적인 변이를 처리하여,행동 인식에 널리 사용되어왔다. 하지만 현재까지 은닉 마르코프 모델을 이용한 행동인식의 방법은 대부분이 손, 머리 또는 인체의 일부분에만 적용하고 있으며, 고차원의 특징 벡터를 가지는 인간 전신의 행동 인식에 적용하기 쉽지 않다는 한계점을 가지고 있다. 특히 기존의 방법은 행동의 시작과 끝을 알고 있는 상태에서 인식을 하고 있다. 이러한 방법은 인식의 성능을 떨어뜨릴 뿐만 아니라, 실제 인식 시스템에 적용하기 위해서는 의미 있는 부분만을 적출(그림 1)하여, 동시에 인식하는 방법이 필요하다. 따라서 본 논문에서는 의미 있는 행동들 간에 발생하는 불필요한 행동은 제거하고, 중요한 행동만을 적출하며, 동시에 인식하는 방법을 제안하였다.

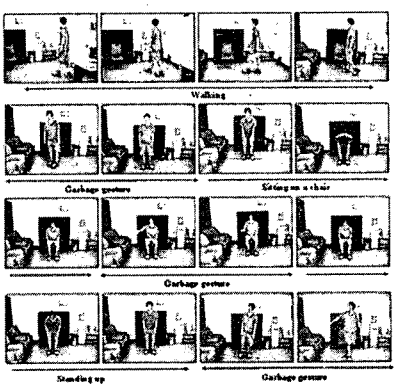


그림 1. 의미있는 제스처 적출의 필요성

## 2. Gaussian Mixture Model을 이용한 특징 그룹화

그룹화 방법을 사용하는 가장 중요한 가정은 서로 다른 동작들은 각각 다른 그룹에 속하게 확률이 높을 것이며, 서로 다른 동작이 같은 그룹에 속한다 할지라도, 해당 그룹에 머무르는 시간과, 그룹의 궤적 변화가 각각의 행동에 따라 크게 다르게 표현된다는데 있다. 그리고 각각의 그룹들은 Gaussian 분포에 의해 잘 표현어 질 수 있다는 것이다. 그룹화를 하기 위해 본 논문에서는 신체의 구성요소와 몸의 중심이 이루는 각을 사용하였으며,

신체의 구성 요소는 왼쪽, 오른쪽 어깨, 팔꿈치, 손목, 골반 뼈, 무릎, 발목 12개를 사용하였다. 그림 2는 걷는 동작, 바닥에 앉는 동작, 바닥에 눕는 동작의 서로 다른 행동에 대해, 12개의 구성요소 중에서 중요 구성요소 3개의 주요 성분의 궤적을 보여주고 있다. 그림 2에서 알 수 있듯이 각각의 동작들은 서로 다른 궤적을 나타내는 것을 알 수 있다.

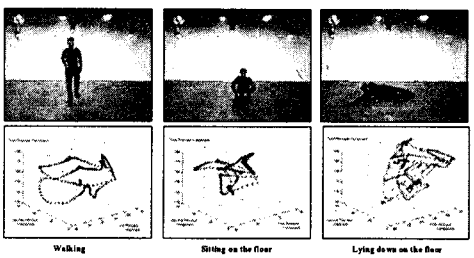


그림 2. 행동들의 서로 다른 궤적

본 논문에서는 데이터들의 Gaussian Mixture Model(GMM)을 예측하기 위해서 Expectation Maximization(EM) 알고리즘을 사용하였다. EM 알고리즘을 이용하여 각각의 Gaussian의 평균과 분산을 계산하였으며, k개의 Gaussian의 집합으로 하나의 그룹을 표현할 수 있게 된다[4]. GMM 파라메터를 예측한 후에 입력되어진 각각의 프레임에서의 동작들과 가장 유사한 Guassian 분포를 가지는 그룹을 계산한다. 이때 유사도가 가장 높은 k번째 그룹의 k가 은닉 마르코프의 출력 문자가 되어진다.

## 3. Garbage 제스처 모델을 이용한 행동 적출 및 인식

### 3.1 Garbage 은닉 마르코프 모델 및 네트워크 구성

본 논문에서는 우리는 2개의 garbage 은닉 마르코프 모델과 1개의 행동 모델을 이용하여, 행동 적출 네트워크를 구성하였다. 1개의 garbage 모델은 의미 있는 행동 앞에 나오는 무의미한 행동을 제거하기 위한 것이며, 나머지 한 개의 garbage 모델은 의미 있는 행동을 결정하기 위한 문턱치 값으로 이용하기 위해 병렬적으로 구성하였다. garbage 제스처 모델을 구성하는 방법은 다음과 같다.

**1:** 출력 함수 $G(b_j(k))$ 은 제스처 은닉 마르코프 모델의 값에 Gaussian 분포를 이용하여, 다시 연산하게 된다.

$$G(\mathbf{b}_{new\_j}(k)) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(b_{old\_j}(k))^2}{2\sigma^2}} \quad (1)$$

여기서 $b_{old\_j}(k)$ 는 학습되어진 행동 은닉마르코프 모델에서 상태 $S_j$에서 k의 문자를 출력할 확률이다.

**2:** 자체 전이 함수의 경우는 학습되어진 제스처 은닉 마르코프 모델의 확률을 그대로 이용한다.

**3:** 상태 전이 확률은 다음과 같이 할당 되어진다.

여기서 $a_{new\_ij}$ 는 garbage 모델의 상태 $S_i$에서 상태 $S_j$로의 전이 확률 함수, $a_{old\_ij}$ 는 학습되어진 제스처 은닉 마르코프 모델의 $S_i$에서 상태 $S_j$로의 전이 확률 함수, 그리고 $N$은 학습되어진 제스처 은닉 마르코프 모델의 전체 상태 계수이다.

$$a_{new\_ij} = \frac{1 - a_{old\_ij}}{N-1}, \quad \text{for all } j, i \neq j. \quad (2)$$

3단계의 수식 (2)의 경우에 모든 상태의 전이가 가능한 ergodic model을 의미하고 있으며, 이렇게 전체 상태 전이가 이루어지므로 의미 없는 행동이 입력되었을 경우 학습되어진 제스처 은닉 마르코프 모델보다 더 높은 확률 값을 가질 수 있게 된다.

이러한 garbage 제스처 은닉 마르코프를 이용한 본 논문에서 제시하는 네트워크는 그림 3과 같다.

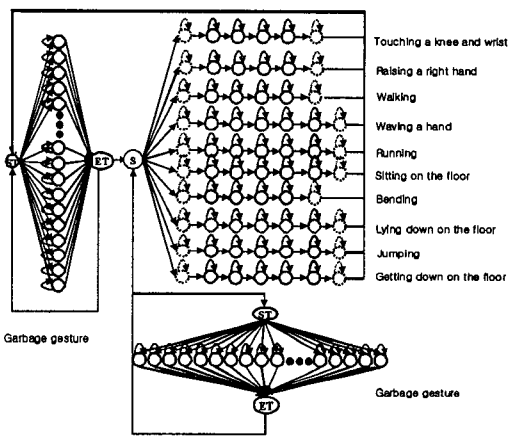


그림 3. 의미 있는 행동 적출 네트워크

### 3.2 의미 있는 행동 적출 및 인식

Garbage 제스처 모델은 의미 있는 행동을 적출하기 위한 문턱치 값으로 이용되어질 수 있다. 즉 입력되어진 행동이 이전에서 설명되어진 은닉 마르코프 네트워크를 통해 계산 되어지는 과정에서, garbage 제스처 은닉 마르코프 모델의 확률 값이 제스처를 표현하는 은닉 마르코프 모델의 확률 값보다 높은 경우에는, 입력 행동이 제거 되어지며 그 이상인 경우에만 적출이 이루어진다. 즉 의미 있는 행동은 오직 garbage 제스처 모델보다 확률 값이 높아질 경우에만 인식되어지면, 다음과 같이 표현된다.

$$P(X \mid \lambda_c) > P(X \mid \lambda_{NG\_model}) \quad (3)$$

여기서 $\lambda_c$ 는 학습되어진 행동 은닉마르코프 모델의 파라미터이며, $\lambda_{NG_model}$은 garbage 제스처 은닉마르코프 모델의 파라미터이다.

## 4. 결과 및 분석

### 4.1 실험 데이터

본 논문에서 제시한 방법을 검증하기 위해 우리는 KU 제스처 데이터베이스[6] 3D motion capture 데이터를

사용하였으며, 행동은 걷기, 뛰기, 점프하기, 인사하기, 바닥에 눕기, 손 흔들기, 바닥에 앉기, 오른손 들기, 바닥에 무릎 꿇기, 허리에 손 올리기의 10가지 행동에 대해 실험하였다. 그러나 이 데이터베이스는 20명에 대해 구성되어 있어, 제시한 방법을 검증하기에 충분히 많은 데이터를 가지고 있지 않기 때문에 우리는 PCA와 Gaussian Random분포를 가지는 계수를 이용하여, 데이터의 양을 확장하였다.

4.2 의미 있는 행동의 적출 결과

KU 제스처 데이터베이스에서 각각의 독립적인 행동을 연결하여, 연속적인 제스처 행동을 생성하였으며, 이를 이용하여 학습과 실험을 하였다. 생성되어진 연속적인 제스처 입력은 하나 이상의 의미 없는 제스처와 의미 있는 제스처로 구성되어 있다.

제안된 방법을 사용하여, 각각의 행동을 나타내는 은닉 마르코프과, garbage 은닉 마르코프 모델의 시간에 따른 확률 값 변화를 그림 4에 표현하였다. 입력 행동의 0초에서 9초까지 구간에서는 garbage 은닉 마르코프 모델의 확률 값이 다른 행동의 은닉 마르코프 모델의 확률 값 보다 높으며, 따라서 이 구간은 의미 없는 구간으로 제거 되어진다. 반면 9초 이후에서는 걷는 행동의 은닉 마르코프 모델의 확률 값이 garbage 은닉 마르코프 모델의 것보다 높으므로, 걷는 행동으로 인식되어진다. 시작점과 끝점을 확인하기 위해서는 Viterbi 알고리즘을 이용하여 찾을 수 있다. 같은 방법으로 20초 이후의 확률 값을 이용하여, 인사하기 행동을 인식할 수 있다.

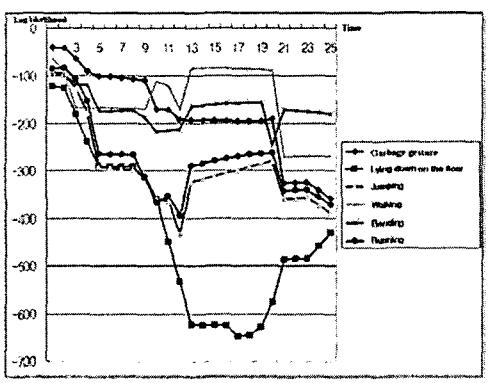


그림 4. 연속동작 입력의 시간에 따른 유사 확률 값

신뢰도를 이용하여, 연속적인 행동에서 적출의 인식률을 평가하였다(표 1). 삽입에러의 경우는 적출기가 존재하지 않는 제스처를 적출하였을 때 발생하는 에러이며, 삭제에러는 적출하여야 할 제스처를 실패하였을 경우에 발생하는 에러이다.

$$\text{신뢰도} = \frac{\text{정확하게 적출되어진행동 개수}}{\text{입력되어진행동개수} + \text{삽입에러개수}}$$

본 논문에서는 모든 행동의 특징들을 고차원의 벡터로 표현하고, 그들 사이에 Gaussian의 분포를 이용한 Gaussian Mixture Model을 이용하여 각 특징을 그룹화 한다. 그리고 각 행동이 그룹을 지나는 궤적을 출력 심볼로 사용하여 은닉 마르코프의 인식에 사용하는 방법을 제시하였다. 이러한 방법론은 연속적인 은닉 마르코프에 비해 계산량이 적으면서도 안정적인 성능을 보여주고 있다. 이 방법은 기존의 손, 머리, 또는 인체의 일부분에만 적용되었던 은닉 마르코프 모델을 확장함으로써 기존의 은닉 마르코프 모델의 방법과 차별성을 가지며, 행동의 시작과 끝을 구분할 수 없었던 형판정합의 문제점을 극복할 수 있었다.

표 1. 의미 있는 행동 적출 결과

| *Gesture* | $N_{IG}$ | $N_{CRG}$ | $N_{DE}$ | $N_{IE}$ | *R* |
|---|---|---|---|---|---|
| 걷기 | 58 | 55 | 1 | 2 | 91.6% |
| 앉기 | 62 | 59 | 0 | 3 | 90.7% |
| 인사하기 | 54 | 54 | 0 | 0 | 100% |
| 점프하기 | 62 | 61 | 0 | 1 | 96.8% |
| 바닥에 눕기 | 61 | 58 | 1 | 2 | 92.0% |
| 손 흔들기 | 60 | 59 | 0 | 1 | 96.7% |
| 바닥에 앉기 | 62 | 58 | 1 | 3 | 89.2% |
| 오른손 들기 | 61 | 62 | 0 | 0 | 100% |
| 바닥에 엎드리기 | 61 | 58 | 1 | 2 | 92.0% |
| 허리에 손 올리기 | 60 | 60 | 0 | 0 | 100% |
| 전체 | 602 | 584 | 6 | 14 | 94.8% |

$N_{IG}$ : 입력되어진 행동 개수
$N_{CRG}$: 정확하게 인식되어진 행동 개수
$N_{DE}$: 삭제에러 개수
$N_{IE}$: 삽입에러 개수
*R*: 신뢰도

## 참고 문헌

[1] A. D. Wilson and A. F. Bobick, "Parametric Hidden Markov Models for Gesture Recognition," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 21, No. 9, pp. 884-900, 1999.

[2] T. Starner, J. Weaver, and A. Pentland, "Real-Time American Sign Language Recognition Using Desk and Wearable Computer Based Video," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 12, pp. 1371-1375, 1998.

[3] R. O. Duda, P. E. Hart, and D. G. Stork, Pattern Classification, *Wiley & Sons*, New York, 2001.

[4] L. R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proc. of IEEE*, Vol. 77, pp. 257-286, 1989.

[5] A. J. Viterbi, "Error Bounds for Convolution Codes and an Asymptotically Optimum Decoding Algorithm," *IEEE Trans. on Information Theory*, Vol. 13, pp. 260-269, 1967.

[6] The KU Gesture Database, *http://gesturedb.korea.ac.kr.*