

# 패턴 성분 기반 인식 범위 축소에 의한 전신 동작 인식

고제인\*, 남양희\*

\*이화여자대학교 디지털미디어학부

e-mail:virginia8002@hanmail.net

## Full-Body Motion Recognition Using Principal Component based Target Reduction

Jane Koh\*, Yang-Hee Nam\*

\*Division of Digital Media, Ewha Womans University

### 요 약

사람의 동작을 인식하는 것에 대한 연구는 게임, 유비쿼터스 컴퓨팅 등의 발전에 따라 그 중요성이 증가하고 있다. 그러나, 대부분의 기존 연구에서는 극히 소수의 동작만을 정의하거나 특정 부위의 동작만을 다루므로 실제 응용에 적용하기에는 적합하지 않다. 본 논문에서는 특정 도메인의 사용 없이, 카메라 영상 입력으로 취득된 동작 패턴 정보만을 이용하여 40종 전신 연속 동작을 구분하는 동작인식 방법을 연구하였다. 인식에 사용된 입력 데이터는 동작자 관절들의 위치 및 회전 값들이며, 다수의 동작들을 인식하기 위해서는 기존의 인식 알고리즘들인 특징기반 인식, HMM, 신경망(Neural Network) 등을 사용하여 복잡한 인식 엔진을 구성하여야 했다. 입력 데이터별로 적합한 인식 모듈을 거치게 하기 위해서는, 동작에 의한 입력 데이터에서 동작자 움직임의 주요 신체 부위를 추출함으로써 입력 데이터가 해당 그룹의 인식 모듈로 자동적으로 분류되게 하는 방법을 사용한다. 이는 다층의 인식 레이어 중 복잡도가 증가하는 하위 레이어일수록 자동 분류에 의해 걸러진 데이터만을 취급하게 되므로 효과적이다. 전체 실험 결과 단계별로 약 79~97%의 인식률을 보였다. 이는 향후 특정 컨텍스트 정보와 결합할 때 매우 높은 인식률을 기대할 수 있게 하는 수칙이다.

### 1. 서론

카메라를 통해 획득된 사람의 움직임 영상에서 동작을 분석하여 활용하는 본격적인 인터페이스에 대한 기대치가 높아지고 있다. 최근 출시된 플레이스테이션 기반의 EyeToy 시스템이나 유비쿼터스 컴퓨팅 환경에 대한 높은 관심도가 그 예이다.

기존의 연구에서는 인식 대상 동작의 수가 10종류 내외로 소수에 그치고 있다. 그 종류가 늘어나는 경우는 인식 대상자의 수를 극소수로 제한하거나 인식을 위한 부가적인 컨텍스트를 활용하여 해석의 모호성을 줄이는 경우들이다. 그러나, 이런 경우에도 인식률은 70~80%대에 머무르고 있다. 따라서, 실험자가 정의하는 실험의 제약을 최소화하는 범용 인식을 위한 연구가 필요하다.

본 연구에서는 마커프리 모션 캡처 장비로부터 신체 움직임에 대한 각 관절부의 3차원 위치 및 각

도 값을 얻어서, 40종 다수의 전신 연속 동작들을 실시간으로 인식하는 방법을 연구하고 실험한다. 이는 다음과 같은 면들에서 의미를 지닌다. 동작 패턴들 간의 구별을 위해 컨텍스트를 사용하지 않고, 신체 관절부의 위치 및 각도 값만을 사용한다. 또한, 전신 연속 동작의 경우 연속문자나 수화 인식과 달리 연결 동작의 일정한 패턴을 얻어내기 힘들며 이를 다룬 연구가 거의 없으므로, 연결 동작 구분 문제를 해결해야 한다.

본 연구에서 인식하려는 대상 동작들은 그 종류가 많고 동작 정의가 다양하다. 그러므로, 입력 데이터를 한꺼번에 처리하지 않고 대상 동작을 대표하는 특성들을 사용하여 자동 분류해줄 방법이 필요하다. 이를 위해 동작의 움직임이 일어나는 주요 움직임 파트를 분석하고 각 파트에 적합한 인식 알고리즘들을 적용하는 방법을 사용한다. 인식 엔진은 서

로 다른 인식 레이어들이 복합되어 다층적으로 구성된다. 각 인식 레이어에서 사용된 알고리즘들은 특징기반인식, HMM, 신경망 알고리즘 등이다.

## 2. 기존 연구

### 2.1 특징 기반 인식

특징 기반 인식은 인식 대상 동작들이 동작자에게서 나타나는 특징들로 충분히 구별 가능한 경우 그 특징을 사용하여 인식하는 방법이다. 가령 3차원 좌표값을 입력 데이터로 다룰 경우, 특정 관절부의 회전 값 및 위치 값을 특정 동작을 나타내는 특징으로 사용할 수 있다. 기존 실험들의 경우, 동작들을 구별할 수 있는 특징을 정의하는 것 자체가 어려운 일이므로 대개 10개 미만인 소수의 동작을 인식 대상으로 제한한 것을 볼 수 있다[1]. 이러한 특징 기반 인식은 대상 동작에서 구분되는 특징을 찾을 수 없을 경우에는 사용할 수 없다는 단점이 있다.

### 2.2 포스처 기반 인식

인식하려는 동작의 과정 중에 특정한 한 자세만으로 동작의 구별이 가능한 경우가 있다. 예를 들어, 본 연구에서 인식하고자 하는 ‘주먹 지르기’ 동작 도중에는 한 발과 다른 쪽의 팔을 앞으로 내민 상태의 자세가 등장한다. 이 자세를 포착함으로써 해당 동작을 인식할 수 있다. 이렇게 특정 순간의 자세로 동작을 인식하는 방법을 포스처 기반 인식이라 한다. 대표적으로 사용되는 알고리즘으로는 정적인 한 상태를 구별하기 위한 신경망이 있다[2]. 그러나 이 방법은 동작의 구별이 시간 흐름에 따른 궤적에 대한 관찰을 통해서만 가능하거나, 동작 도중의 상태들이 거의 비슷할 경우에는 사용될 수가 없다.

### 2.3 제스처 기반 인식

특정 프레임에서의 자세로 구별이 어려워 동작이 발생하는 동안 나타나는 특정 관절부의 궤적을 통하여 인식을 수행할 경우, 이를 제스처 기반 인식이라 한다. 사용되는 대표적 알고리즘으로 HMM(Hidden Markov Model), TDNN(Time Delayed Neural Networks), RNN등이 있다. 기존 연구 중에는 대상 동작을 목표 물체와 연관시켜 동작자의 시선과 목표 물체를 컨텍스트 정보로 사용한 예가 있다[3]. 이러한 방법은 관절의 궤적 정보만을 사용하는 경우보다 인식률의 향상을 보인다.

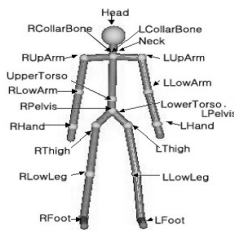
본 논문에서 하고자 하는 40종 동작의 인식은 그 대상동작 정의의 분포가 다양하여 우선, 주요 신체

움직임 파트를 추출하는 과정이 필요하다. 그리고, 분류된 각 파트에 따라 인식에 적합하므로 다른 알고리즘들을 적용시키는 방법을 택한다.

## 3. 패턴 성분 기반 전신 동작 인식 인식

### 3.1 동작자 관절 및 입력 데이터

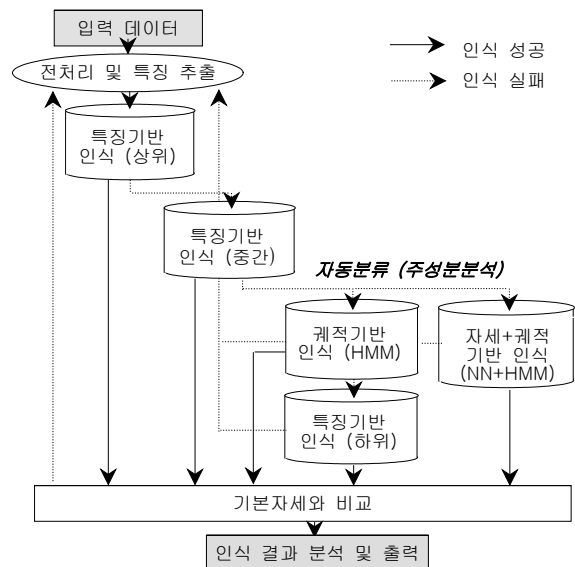
동작자의 관절은 [그림 1]과 같이 총 20개로 이루어져 있다. 이러한 20개의 관절은 4개 관절씩 구성된 몸통, 오른/왼쪽 팔, 오른/왼쪽 다리의 5가지 주요 신체 파트로 구분될 수 있다. 동작자가 연속으로 동작을 취함에 따라, 20개 각 관절로부터 위치와 각도의 총 6차원 값들이 병렬적으로 입력된다.



[그림 1] 동작자의 관절

### 3.2 패턴 성분 기반 인식 시스템 구조

시스템의 전체 구조는 [그림 2]에서 보여지며, 인식기 내부는 다수의 인식 모듈들로 구성된다. 이는 40종 대상 동작들의 정의가 다양하므로, 대상 동작의 주요 움직임 파트를 분석하고 그에 따라 적합한 인식 알고리즘을 적용하는 방법을 사용하기 때문이다. 본 논문에서는 주요 움직임 파트를 분석함으로써 해당 그룹의 인식 모듈로 자동 분류하기 위해, 차원축소를 통하여 저차원상에서 변수의 관계를 규명하는 다변량 자료분석기법인 주성분분석(principal component analysis) 방법을 사용한다.



[그림 2] 시스템의 전체 구조도

본 논문에서는 20개 관절의 위치와 회전 각각 60 차원 변수들을 포함하는 행렬  $X$ 이 생성된다.

$$X \sim (\mu, \Sigma)$$

$$\text{, where } X = \begin{bmatrix} X_1 \\ X_2 \\ \dots \\ X_p \end{bmatrix}, \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \dots \\ \mu_p \end{bmatrix}, \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2p} \\ \dots & \dots & \dots & \dots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp} \end{bmatrix}$$

행렬  $X$ 의 일반적인 특징은 분산과 공분산  $(\mu, \Sigma)$ 에 내재되어 있으나 다변량의 구조나 특성을 파악하는 데는 충분치 못하다. 따라서, 공분산 행렬에 대한 고유값과 고유벡터가 각각

$$(\lambda_1, e_1), \dots, (\lambda_p, e_p), \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$$

일 때,  $i$  번째 주성분

$$Y_i = e_i' X, \quad i = 1, 2, \dots, p$$

을 찾는다. 이처럼 주성분을 찾는 문제는 공분산 행렬의 고유값과 고유벡터를 구함으로써 구해지며, 구해진 주성분은 전체 입력 데이터에 대한 차원 축소를 통해 주요 움직임 파트만을 구별해 보여준다.

시스템의 구조는 우선 데이터의 잡음 제거를 위한 averaging 방법이 전처리 과정에서 이뤄진 후, 각각 몸통과 다리 중심의 특징을 통해 동작들을 추출하기 위한 특징 기반 인식 상위와 중간 레이어를 거친다. 아직 인식되지 못한 경우 데이터 처리의 복잡성을 줄이기 위하여, 주요 움직임 파트 분석을 통해 파트별로 적합한 레이어를 거치게끔 분류하는 과정을 필요로 한다. 이때 주성분분석을 통해 팔 부위가 주요 움직임 파트로 분석된 경우 손의 궤적을 사용하는 궤적 기반 인식 레이어와 팔 부위 특징을 사용하는 특징기반 인식 하위 레이어를 차례로 거친다. 반면, 다리 부위가 주요 파트로 분석되는 경우 자세 및 궤적을 동시에 기반으로 사용하는 레이어로 분류된다. 이 같은 방법을 통해 해당 그룹 레이어로 분류하는 것은, 하위로 갈수록 복잡도가 커지는 인식 레이어들이 자동으로 분류된 데이터들만을 다루게 함으로써 효율성을 높인다.

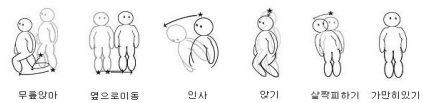
연속 동작 인식 시 하나의 동작 패턴이 더 큰 동작 패턴의 부분에 포함되거나 연결 동작이 의미 있는 동작으로 오인식되는 경우가 있다. 이는 기본 자세가 등장할 때까지 인식의 최종 결정을 지연시킴으로써 오인식을 판별하는 방법을 사용한다.

### 3.3 부분별 인식 알고리즘

#### 3.3.1 특징 기반 인식

특징 기반 인식은 인식의 특징으로 사용되는 데이

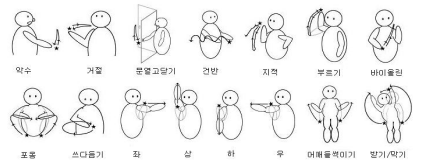
터의 종류에 따라 세 가지로 나뉘는데, 상위 레이어는 몸의 자세나 무게중심의 방향, 중간 레이어는 다리의 자세나 이동 방향 등을 특징으로 사용한다. 이들은 상대적으로 그 복잡도가 적은 인식 모듈들로 주성분분석이 이루어지기 전 매 프레임마다 인식을 수행한다. 이에 비해 상대적으로 복잡성이 큰 특징기반 하위 레이어에서는 팔의 자세나 위치 성분을 특징을 사용한다. 이는 팔 중심 동작으로 분류된 데이터들에 대해서 궤적 기반 레이어에서 인식을 못한 경우에만 인식을 수행한다. 레이어별 인식 대상 동작들은 각각 [그림 3,4,5]의 동작들이다.



[그림 3] 상위 레이어 인식 대상 동작



[그림 4] 중간 레이어 인식 대상 동작



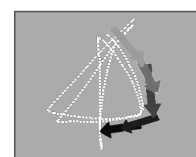
[그림 5] 하위 레이어 인식 대상 동작

#### 3.3.2 HMM을 이용한 궤적 기반 인식

주성분분석으로 팔 부위만이 주된 움직임을 보이는 부분으로 판별된 경우에는, 먼저 시간 흐름에 따라 변화하는 팔에 의한 궤적 패턴에 의미가 주어지는 동작들을 인식하기 위해 궤적 기반 인식 레이어를 적용한다. 본 논문에서는 [그림 7]에서처럼 특정 순간의 정보로는 인식이 불가능한 궤적 발생 동작에 대해 현재 시점으로부터 시간을 거슬러 올라가며 의미 있는 동작을 spotting하는 HMM의 백워드 알고리즘 및 포워드 알고리즘, 훈련 알고리즘으로는 Baum-Welch 알고리즘을 사용한다[4]. 해당 인식 대상 동작들은 [그림 6]에 보여진다.



[그림 6] 궤적 기반 인식 대상 동작



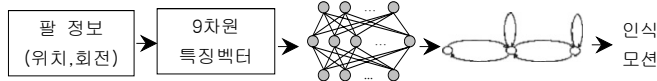
[그림 7] 연속 '세모' 동작 인식

### 3.3.3 NN+HMM을 이용한 자세+궤적 기반 인식

주성분분석이 팔 부위만이 아닌 다리를 함께 움직임의 주요 파트로 판별하는 경우, 팔과 다리가 동시에 사용된 특정 순간의 자세를 통해 인식하는 방법을 사용한다. 이를 위해서는 신경망의 출력 시퀀스를 HMM의 입력으로 사용하여 두 알고리즘을 복합 활용하는 방법을 사용하는 것이 유용하다[그림 9]. 이는 신경망이 통계적인 방법으로 정지된 한 순간의 모양이나 특징을 구별하기에 적합하며, HMM은 동작 시퀀스에 대한 시간에 따른 통계적 인식에 적합하기 때문이다. [그림 8]에 해당 대상 동작들이 보여진다.



[그림 8] 자세+궤적 기반 인식 대상 동작



[그림 9] 자세+궤적 기반 인식 모듈 구조

## 4. 동작 인식기 구현 및 결과

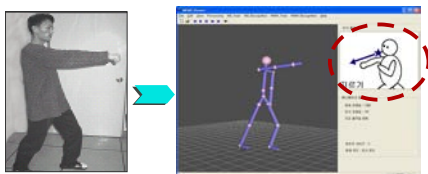
### 4.1 HMM 구현

본 논문에서는 back-loop 없이 좌에서 우 방향으로만 천이되는 Left-to-Right HMM을 사용한다. HMM에 입력될 심볼들의 제공을 위해서는 먼저, 입력되는 3차원의 궤적 데이터로부터 2차원 fitting plane을 Least Square Method로 추출한다[5]. 그 평면상의 방향 벡터들로부터 가장 안정적인 심볼들을 제공하는 것으로 실험된 8방향의 체인 코드를 구현 후, 이를 HMM의 입력 시퀀스로 사용한다.

### 4.2 NN+HMM 구현

신경망은 입력 데이터로부터 팔 동작의 9차원 특징 벡터가 추출되어 9개 입력노드로 설정된다. 은닉층은 한개, 출력노드는 5개로 '주먹지르기', '주먹휘두르기' 동작 각각의 시작과 끝, '상단막기' 동작의 끝 등에 해당된다. 신경망의 5개 출력노드는 HMM의 5레벨로 양자화 된 입력 시퀀스로써 적용된다.

### 4.3 구현 결과



[그림 10] "주먹 지르기" 동작과 인식되는 화면

구분	인식률(%)
궤적 기반 인식	97.90
자세+궤적 기반 인식	83.50
특정 기반 인식 (상위)	87.64
특정 기반 인식 (중간)	78.72
특정 기반 인식 (하위)	82.46
합계	86.04

[표 1] 전신 연속 동작 인식률

실험 결과 전체적으로 약 86%의 인식률을 보였다 [표 1]. 이와 별도로 오인식 분석을 행한 결과, 몇몇 동작의 쌍들에 대해 오인식이 발생하는 현상을 볼 수 있었다. 이러한 현상은 향후 컨텍스트 정보를 활용함으로써 해결할 수 있을 것이다.

이외에도, 제스처 기반 인식 방법 중 하나인 TDNN을 적용하여 11개의 팔 타입 동작에 대해 실험해 보았다. 실험결과 약80~85% 정도의 인식률을 보였으나 실시간으로 적용하기에는 너무 느려서 채택하지 않았다.

## 5. 결론 및 향후 과제

본 논문에서는 동작 패턴 정보만을 이용하여 다수의 동작을 인식함으로써 일반화된 인식엔진 가능성을 실험하였다는데 의의가 있다.

향후 인식 엔진의 성능을 향상시키기 위해서는 컨텍스트의 사용을 개발된 엔진과 통합하고, 동작에 따른 개인별 차이의 정도를 인식하여 동작의 종류를 설정하는 기법의 도입이 유용할 것이다.

### 참고문헌

- [1] Anjum Ali and J. K. Aggarwal. "Segmentation and Recognition of Continuous Human Activity", IEEE Workshop on Detection and Recognition of Events in Video, pp.28-36, Vancouver, Canada, July 2001
- [2] R. Beale and A. Edwards. "Recognition postures and gestures using neural networks.", In R. Beale and J. Finlay, editors, Neural Networks and Pattern Recognition in Human Computer Interaction. Ellis Horwood, 1992
- [3] Kaisa Vaananen and Klaus Bohm "Gesture driven interaction as a human factor in virtual environments an approach with neural networks". In R. Earnshaw, M. Gigante, and H. Jones, editors, Virtual Reality Systems, pp. 93-106. Academic Press, 1993
- [4] R. Dugad and U. B. Desai. "A Tutorial on Hidden Markov Models", Indian Institute of Technology Report. 1996[5] Y. Nam, K. Wohn. "Recognition of Hand Gestures with 3D, nonlinear arm movements", Pattern Recognition Letters. 1997