

# 은닉 마르코프 모델 기반 동작 인식 방법

김종호<sup>†</sup>, 윤요섭<sup>\*\*</sup>, 김태영<sup>\*\*\*</sup>, 임철수<sup>\*\*\*\*</sup>

## 요 약

본 논문은 비전 기반 동작 인식 방법으로 모범 동작의 유형을 모형화하고 이를 이용하여 사용자의 동작을 인식하고 모범동작과 사용자의 동작간의 유사도를 측정하는 방법을 제안한다. 동작 인식을 위하여 은닉 마르코프 모델 기반의 유형화 기법을 통하여 모범 동작의 유형 모델을 구성하고 이를 이용하여 사용자의 동작을 인식한다. 유사도 측정을 위하여 편집 거리 알고리즘을 응용하여 모범 동작과 사용자 동작의 유사도를 측정하고 점수 표기가 가능하도록 하였다. 본 논문에서 제안하는 동작 인식 처리 방법은 평균 93% 이상의 높은 인식율을 보였다. 본 연구의 결과는 동작 인식 기반 게임, 자세인식, 동작의 반복 훈련 및 훈련 달성도 측정을 요하는 재활훈련 시스템 등에 활용 가능하다.

## Human Primitive Motion Recognition Based on the Hidden Markov Models

Jong-Ho Kim<sup>†</sup>, Yo-Seop Yun<sup>\*\*</sup>, Tae-Young Kim<sup>\*\*\*</sup>, Cheol-Su Lim<sup>\*\*\*\*</sup>

## ABSTRACT

In this paper, we present a vision-based human primitive motion recognition method. It models the reference motion patterns, recognizes a user's motion, and measures the similarity between the reference action and the user's one. In order to recognize a motion, we provide a pattern modeling method based on the Hidden Markov Models. In addition, we provide a similarity measurement method between the reference motion and the user's one using the editing distance algorithm. Experimental results show that the recognition rate of ours is above 93%. Our method can be used in the motion recognizable games, the motion recognizable postures, and the rehabilitation training systems.

**Key words:** Motion Recognition(동작 인식), Similarity Measurement(유사도 측정), Hidden Markov Models(은닉 마르코프 모델)

## 1. 서 론

최근 유비쿼터스 기술과 가상현실 기반 기술들이 발전함에 따라 동작 인식 처리 기술들이 많이 연구되고 있다[1,2]. 이러한 동작 인식 기술을 기반으로 하는 연구들은 손동작을 이용한 전자기기와 컴퓨터 제

어에 관한 HCI(Human Computer Interaction)연구, 사람에게 부착한 센서를 통해 전달되는 데이터를 이용해 움직임을 인식하여 컴퓨터 애니메이션 등에 활용하는 연구[1]등이 있다. 또한, 카메라를 통해 입력되는 영상을 분석하여 동작을 인식하거나 추적하는 시스템 연구[2], 보안이나 감시를 요하는 장소에서

※ 교신저자(Corresponding Author): 김태영, 주소: 서울 성북구 정릉동 16-1(136-704), 전화: 02)913-1362, FAX: 02)913-1362, E-mail: sebhi@naver.com

접수일: 2008년 12월 1일, 완료일: 2009년 2월 17일

<sup>†</sup> 서경대학교 컴퓨터 공학과 석사 과정

(E-mail: sebhi@naver.com)

<sup>\*\*</sup> 준회원, 서경대학교 컴퓨터 공학과 석사 과정

(E-mail: yysddt@hanmail.net)

<sup>\*\*\*</sup> 정회원, 서경대학교 컴퓨터 공학과 조교수

<sup>\*\*\*\*</sup> 정회원, 서경대학교 컴퓨터 공학과 부교수

(E-mail: cslim@skuniv.ac.kr)

※ 본 연구는 문화관광부 및 한국문화콘텐츠진흥원의 문화콘텐츠기술연구소(CT)육성사업의 연구결과로 수행되었음.

위험한 동작들을 자동적으로 찾아내는 감시 시스템 등 다양한 기술에의 응용이 가능한 분야이다.

이러한 동작 인식에 대한 연구들은 동작을 인식하기 위해 연속된 프레임에서 사용자의 신체 특징 추출과 같은 많은 계산이 필요하고 추출된 특징 벡터에서 작업의 패턴을 인식하는 과정이 필요하다. 동작의 패턴 인식은 주로 제스처 인식을 사용하는데 제스처 인식은 인체 각 부위가 시간 축에 대해 어떠한 형상 변화를 가지는가를 자동으로 알아내는 것을 의미한다. 그러나 인체는 매우 복잡한 3차원 관절구조를 지니고 있어서 자동으로 제스처를 인식하는 것은 매우 어렵다. 따라서 제스처 인식은 인간의 손, 팔 등이 기계에 의해서 측정될 수 있어야 한다는 점을 요구한다[3,4].

손의 제스처나 몸의 움직임 같은 사람의 구체적 동작을 인식하기 위해 컴퓨터 비전 분야에서는 비디오 시퀀스 영상에 대한 연구가 활발하게 진행 중이다. 최근에는 시간의 변화에 따른 영상의 특징점 추출 정보를 이용하는 은닉 마르코프 모델이나 모션벡터를 이용하여 제스처를 측정하려는 시도가 많이 이용되고 있다[4,5].

인간의 동작을 인식하는 시스템은 여러 응용분야에서 다양하게 이용될 수 있는 분야로 게임이나 오락 분야에서는 사용자들이 자신의 신체를 이용하여 가상의 기구나 장비들을 제어할 수 있으며, 운동 자체에 의미를 두는 자세 교정이나 재활 훈련과 같은 치료 분야에서도 이용될 수 있다.

본 논문에서는 2차원 영상을 통한 동작 인식 방법으로서 전처리 과정에서 여러 정규 동작들에 대한 데이터를 준비하고 이를 이용하여 사용자 동작을 비교한다. 정규 동작들과 사용자 동작은 은닉 마르코프 모델(HMM:Hidden Markov Model) 기반 패턴 인식 방법[6-8]을 통하여 비교하였고 동작의 유사도 측정 방법으로는 편집 거리 알고리즘(Edit Distance) [9-12]을 사용하였다. 기존의 비전 기반 동작 인식 방법들을 2장에서 소개하고, 3장에서 은닉 마르코프 모델 기반 동작 인식 방법을 설명한다. 4장에서는 동작에 대한 평가 기준으로서 점수화를 위한 유사도 측정 방법을 설명하고 5장에서 동작 인식과 유사도 측정에 대한 실험과 결과를 나타낸 뒤 마지막으로 6장에서 연구에 대한 결론과 향후 연구 방향에 대하여 논한다.

## 2. 기존 비전 기반 동작 인식 방법 및 최근 연구 동향

### 2.1 3차원 모델 기반의 센서와 마커를 이용한 동작 인식

이 방법은 인간의 신체에 센서를 부착하여 움직임 정보를 얻어낸 후 위치 데이터를 컴퓨터 내부에 모델링된 캐릭터에 적용하여 유사한 동작을 하도록 하거나 동작에 관한 데이터를 분석하여 인식에 이용하는 방법이다. 자연스러운 동작 데이터를 비교적 빠른 시간에 손쉽게 얻을 수 있고, 일단 획득한 동작 데이터는 데이터베이스화 하여 나중에 재사용 하는 것이 가능하므로, 전 세계적으로 급성장하고 있다. 위치와 방향을 측정하는 방법으로는 기계적 방식, 자기 방식, 광학 방식의 인식 장비들을 많이 사용하고 있는데, 이러한 장비들은 대부분 고가이며 제한적 환경에서만 응용이 가능하다.

### 2.2 시각 기반의 영상정보를 이용한 동작 인식

비전 기반의 영상 정보를 이용한 동작 인식 방법은 비접촉식 인식 방법으로 인간의 신체에 어떤 장치도 부착하지 않고 카메라를 통해 얻어진 영상을 이용하여 외관의 특징을 기반으로 제스처를 인식한다[6, 13-15]. 동작에 대한 분류나 구분을 하기 위해서 동작성에 대해 모델들과의 비교가 이루어지는데 여기에서는 다양한 특징 정보들이 이용된다. 이러한 특징에는 기하학적 특징, 통계적 특징, 시간변화에서 추출된 특징 등이 이용된다.

첫 번째 기하학적 특징을 이용하는 방법은 경계, 코너 등을 구하여 정규 모델에 대한 데이터를 준비하고 인식할 동작에 대해서도 같은 방식으로 구한 데이터와 비교하는 방법[13]이다. 기하학적 특징을 이용하면 여러 가지 복잡한 동작에 대한 특징 추출이 어렵고 많은 전처리 과정이 필요하게 되어 인식 속도와 인식률이 떨어지게 된다.

두 번째 통계적 특징을 이용하는 방법은 전처리 과정으로 많은 영상 집합에 대해 통계적으로 높은 확률 값을 가지는 밝기 정보와 위치 정보를 분석하여 비교하는 방법[6]이다. 비교적 좋은 인식률을 가지며 실시간 적용에도 효과적이다.

세 번째로 시간의 변화에 따른 영상의 특징점 추출 정보를 이용하는 방법은 동작이 이루어지는 대상

에 대해 시간의 흐름에 따른 동작의 패턴을 분석하여 패턴에 따른 동작을 구분하거나 분류하는 방법 [14-15]이다. 이러한 방법은 일반적인 환경에서도 동작 성을 쉽게 추출할 수 있는 장점이 있지만 수많은 영상들을 이용하여 학습시키고 적당한 패턴을 모델화 하여야하는 어려움이 있다.

### 2.3 최근 연구 동향

동작 인식과 관련된 연구동향을 살펴보면 특히, 손동작 인식과 그의 응용인 수화 인식에 대한 연구 [16-17]와 표정인식[18] 연구가 활발하게 진행되고 있다. 이러한 연구들에서는 동작 인식이 손이나 얼굴 등 신체의 일부분을 대상으로 특징점을 분석하여 인식하기 때문에 동작 인식 부위에 대하여 제한적이라는 단점이 있다. 전신 동작에 대한 인식 연구로는 동작의 외관 정보를 이용하여 인식하는 연구[19]와 움직임 정보(Motion History)를 이용한 연구[20] 등이 있다. 이러한 방법들은 시간축상 움직임 정보를 하나의 이미지로 만들어 표현함으로써 복잡한 모델링 없이 좋은 인식 결과를 얻을 수 있는 장점이 있으나, 인식 기준이 되는 템플릿과 인식 대상 동영상 간의 시·공간적 제약성이 높고 실루엣 등의 외형적 의존도가 높기 때문에 주변 환경에 따라 영향을 받는 한계가 있다[21].

본 논문은 전신 동작을 분석하기 위해 특징점을 신체의 대표부위에 제한적으로 부여 하였으며 동작의 움직임 정보 추출을 위하여 고가의 센서 장비를 사용하는 대신 적외선 발광 소자(LED)와 영상 처리를 통한 필터기법을 이용하여 신체 일부에 대한 특징점의 움직임 정보를 획득한다. 이러한 움직임 정보는 시간 변화에 따른 움직임 정보로서 은닉 마르코프 모델을 사용하여 패턴을 모델화하고 모델 정보와 사용자 동작간의 동작을 비교하여 유사여부를 판단한다. 또한 은닉 마르코프 모델 기법이 신체 일부에 제한적으로 적용되었던 것에 비하여 본 논문에서 제안하는 모델링 기법은 전신 부위에 대한 움직임 패턴 분석에 용이하고 움직임 정보 분석 방법에 있어서 시작 시각을 자동화하여 처리한다.

## 3. 동작 인식

본 논문에서는 첫 번째 과정으로 팔, 다리와 같은

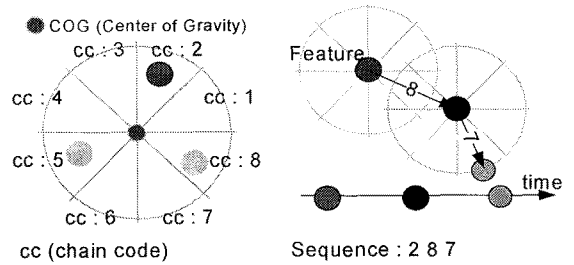


그림 1. 임의의 특징점에 대한 무게중심 기준 구간 기호화(좌) / 시간 변화 기준 구간 기호화 방법(우)

인체 부위 특징점을 추출하고 이를 이용한 패턴 인식을 통하여 동작을 인식한다. 인체 부위 특징점 추출 방법으로는 적외선 발광 소자를 이용하여 머리, 양손목, 양 발목의 5개 부분에 대하여 위치 좌표를 얻어낸 후, 각 부분에 대한 시간변화에 따른 위치 변화를 8개 구간으로 분류하여 사용하였다. 그림 1은 임의의 특징점의 위치와 시간에 따른 위치 변화를 기호화하는 방법을 나타낸다. 방향 벡터는 8구간으로 구분된 공간에 배치되고, 시간에 따른 운동 벡터가 포함되는 구간이 해당 벡터의 기호가 된다. 여기에서 기호는 계산 편의를 위하여 1~8까지의 숫자로 나타내었다. 이 방법만으로는 처음 시작 위치에 대한 기호가 생성되지 않기 때문에 처음 시작 위치에 대한 기호는 모든 특징점들을 ‘Convex Hull’ 기법을 이용하여 다각형으로 만든 후에 무게 중심을 구하고, 무게 중심점을 원점으로 하는 공간에 재배치하여 해당 특징점이 속하는 구간의 기호가 시작 위치에 대한 기호가 되도록 하였다. 이러한 방법에 의해 계산된 시간 변화에 따른 위치변화 집합(Sequence)은 동작에 대한 패턴 정보로 사용할 수 있다.

### 3.1 동작 패턴 모델링

동작을 인식하기 위해서는 동작 고유의 데이터를 패턴화하고 이러한 패턴 데이터를 사용목적에 맞게 구조화하는 모델링 과정이 필요하다. 이러한 패턴 모델링 기법은 보통 확률적 접근 방법이 일반적이다. 본 논문에서는 패턴의 모델링 방법으로서 음성인식, 손동작 인식 등에 사용되는 은닉 마르코프 모델 기법을 사용하였다.

은닉 마르코프 모델은 모델링하기 어려운 패턴의 확률론적 구조를 명확하게 표현 가능한 기호들을 통하여 모델링하는 이중 확률론적 구조의 모델이다. 이

러한 은닉 마르코프 모델은 패턴을 유한 상태 오토마타(Probabilistic finite state automata) 구조로 표현하며 패턴의 유한 상태들은 시간 전후에 따른 결합 확률(Joint probability)로 나타낼 수 있다. 따라서 모델링을 위한 기호 집합은 패턴의 고유 특징과 시간 변화에 대하여 전후 관계성을 포함하는 형식이 되어야 한다[7]. 본 논문에서는 이러한 고유 데이터-모델링 가능한 기호들-로 시간 변화에 따른 위치 변화 집합을 사용하였다.

먼저 패턴 데이터와 시간과의 관계를 표현하기 위하여 패턴의 유한 상태들을 시간  $t$ 에 대하여 나타내고, 행동을 완료하는 순간  $t_n$ 까지 전개되는 과정을 모델링하기 위하여  $t+1$ 의 상태를  $t$  상태의 영향력으로 표현한다.  $t$ 에서의 상태를  $w(t)$ 라 하면, 전체 길이  $T$ 의 상태 시퀀스에서의 특정 상태  $w_i$ 는 식 1과 같이 표현 가능하며, 시간  $t$ 일 때 상태  $w_i(t)$ 에서 시간  $t+1$ 일 때 상태  $w_j(t+1)$ 가 될 확률은 식 2와 같이 전이 확률(Transition Probability)  $a_{ij}$ 로 표현 할 수 있다.  $a_{ij}$ 의 전체 집합 모델  $\theta$ 와 전체 길이  $T$ 의 상태 시퀀스가 있다면, 모델  $\theta$ 가 이 상태 시퀀스를 생성할 확률은 일련의 상태들 간의 전이 확률을 곱하는 것으로 나타낼 수 있다.

$$w_i = \{w(1), w(2), \dots, w(T)\} \quad \text{식 (1)}$$

$$P(w_j(t+1)|w_i(t)) = a_{ij} \quad \text{식 (2)}$$

표현 가능한 기호들을 통하여 패턴을 모델링하는 방법은 특정 상태에서 발생 가능한 기호들의 확률을 나타내는 것으로 접근한다. 시간  $t$ 에서의 기호를  $v(t)$ 라 하면, 전체 길이  $T$ 의 기호 시퀀스에서 특정 기호  $v_k$ 는 식3과 같이 표현 가능하며, 상태  $w_j$ 에서 관측될 수 있는 기호-동작의 운동 벡터 정보-를  $v_k$ 라 하면, 시간  $t$ 의 어떤 상태  $w_j(t)$ 에서 기호  $v_k(t)$ 가 발생할 확률은 식 4와 같이 기호 관측 확률  $b_{jk}$ 로 표현할 수 있다. 즉, 전체 길이  $T$ 에 대한 기호집합이 주어졌을 때 시간  $t$ 에 대한 기호  $v_k(t)$ 가 발생할 확률이 높은 상태  $w_j(t)$ 를 찾아감으로서 전체 상태 시퀀스를 모델링할 수 있다.

$$v_i = \{v(1), v(2), \dots, v(T)\} \quad \text{식 (3)}$$

$$P(v_k(t)|w_j(t)) = b_{jk} \quad \text{식 (4)}$$

본 논문에서는 모델의 상태와 기호의 개수를 각각

10, 8로 설정하였으며 이에 따른 상태 전이 확률과 각 상태에서의 기호 발생 확률을 정리하면 다음과 같다.

$$0 \leq t < T$$

$$W^T = \{w|w_n, 0 \leq n < c\}, V^T = \{v|v_n, 0 \leq n < k\}$$

$c$  : number of states,  $k$  : number of symbols

$$A = \{a_{ij} = P(w_j(t+1)|w_i(t)) | w_i \in W^T, w_j \in W^T\}$$

$$B = \{b_{jk} = P(v_k(t)|w_j(t)) | v_k \in V^T, w_j \in W^T\}$$

식 (5)

$$\sum_j a_{ij} = 1 \text{ for all } i$$

$$\sum_k b_{jk} = 1 \text{ for all } j$$

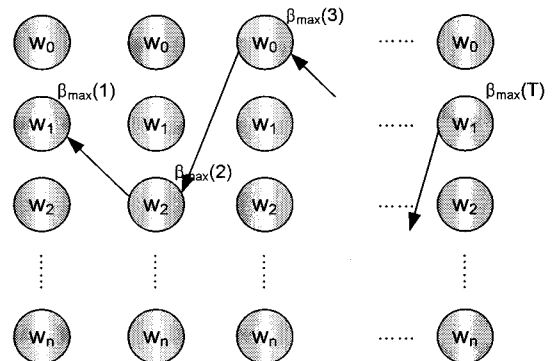
정리된 식 5에 기반 하여 동작 모델  $\theta$ 에 대한 파라미터 집합 A와 집합 B의 값은 모델링할 동작의 기호 시퀀스를 이용한 학습과정을 통하여 구할 수 있다. 파라미터 값 계산은 다음과 같은 역방향 알고리즘[8]을 사용한다. 모델  $\theta$ 가 상태  $w_i(t)$ 에 있을 때  $t$ 까지의 상태 시퀀스 생성 확률을  $\alpha_i(t)$ 라하고  $t+1 \rightarrow T$ 까지의 나머지 상태 시퀀스 생성확률을  $\beta_i(t)$ 라 한다면 다음과 같이 표현할 수 있다.

$$\beta_i(t) = \begin{cases} 0 & w_i(t) \neq w_0, t = T \\ 1 & w_i(t) = w_0, t = T \\ \sum_j \beta_j(t+1) a_{ij} b_{jk} v(t+1) & \text{otherwise} \end{cases} \quad \text{식 (6)}$$

식 6을 이용하여  $\beta_i(T-1)$ 를 구하면 식 7이 된다.

$$\beta_i(T-1) = \sum_j a_{ij} b_{jk} v(T) \beta_j(T) \quad \text{식 (7)}$$

구해진  $\beta_i(T-1)$ 를 이용하여  $\beta_i(T-2)$ 를 구할 수 있으며, 이와 같은 과정을 반복해서  $\beta_i(t)$ 를 구할 수 있다.



본 논문에서는 모델의 상태와 기호의 개수를 각각

식7에서 결정되어야 하는 전이 확률  $a$ 와 관측 확률  $b$ 의 추정 값을 학습을 위한 기호 시퀀스  $V^T$ 를 통하여 구할 수 있다. 상태  $w_i(t-1)$ 에서  $w_j(t)$ 로의 전이 확률을  $\gamma_{ij}(t)$ 라 하면 식 8과 같이 나타낼 수 있다.

$$\gamma_{ij}(t) = \frac{\beta(t-1)a_{ij}b_{jk}\beta_j(t)}{P(V^T | \theta)} \quad \text{식 (8)}$$

상태  $w_i(t-1)$ 에서  $w_j(t)$ 로의 전이 기대 수는  $\sum_{t=1}^T \gamma_{ik}(t)$ 로 나타낼 수 있으며,  $w_i$ 로부터의 임의의 전이 전체 기대 수는  $\sum_{t=1}^T \sum_k \gamma_{ik}(t)$ 로 나타낼 수 있다. 따라서  $\hat{a}_{ij}$ 는  $w_i$ 에서  $w_j$ 로의 전이 기대 수와  $w_i$ 로부터의 임의의 전이의 대한 전체 기대 수간의 비율로 찾아낼 수 있으며 같은 방법으로 상태  $w_j$ 에서 특정  $v_k$ 가 관측되는 빈도와 임의의 기호에 대한 빈도 간의 비율을 계산하여 추정  $\hat{b}_{jk}$ 를 구할 수 있다.

$$\hat{a}_{ij} = \frac{\sum_{t=1}^T \gamma_{ik}(t)}{\sum_{t=1}^T \sum_k \gamma_{ik}(t)} \quad \text{식 (9)}$$

$$\hat{b}_{jk} = \frac{\sum_{t=1}^T \sum_l \gamma_{jl}(t)}{\sum_{t=1}^T \sum_{l=v_k} \gamma_{jl}(t)} \quad \text{식 (10)}$$

식 9와 식 10을 이용하여 다음과 같은 알고리즘을 통해 필요한 모델 파라미터 집합 A와 집합 B를 구하였다.

```

ε ← 0.00001
begin Training
: Initialize  $a_{ij}, b_{jk}$ , training sequence  $V^T, z \leftarrow T$ 
while  $\langle i,j,k \rangle \max [a_{ij}(z) - a_{ij}(z-1), b_{jk}(z) - b_{jk}(z-1)] < \epsilon$ 
    compute  $a(z)$  from  $a(z-1)$  and  $b(z-1)$  by Eq.7
    compute  $b(z)$  from  $a(z-1)$  and  $b(z-1)$  by Eq.8
     $a_{ij}(z) \leftarrow \hat{a}_{ij}(z)$ 
     $b_{jk}(z) \leftarrow \hat{b}_{jk}(z)$ 
     $z \leftarrow z - 1$ 
endwhile
return  $a_{ij} \leftarrow a_{ij}(z); b_{jk} \leftarrow b_{jk}(z)$ 
end Training
    
```

### 3.2 동작 패턴 매칭 방법

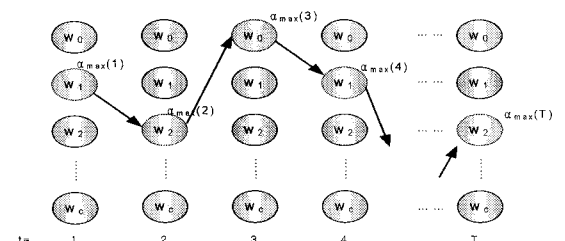
모델  $\theta$ 는 구해진 파라미터 집합 A와 집합 B로 나

타낼 수 있다.  $b_{jk}$ 는 어떤 상태  $w_j(t)$ 에서의 기호  $v_k(t)$ 의 발생 가능성이고 상태  $w_i(t-1)$ 에서 다음 상태  $w_j(t)$ 로의 전이 확률  $a_{ij}$  결정과 밀접한 관계를 가진다. 즉, 모델  $\theta$ 는 학습에 사용된 기호 시퀀스에 최적화된 파라미터 값을 가진다고 말할 수 있다. 본 논문에서는 동작을 인식하기 위해서 학습을 통해 구해진 모델  $\theta$ 에 대하여 사용자의 행동에 따른 기호 시퀀스와 학습에 이용하였던 기호 시퀀스를 이용하여 각각의 매칭 확률을 구하고 두 확률의 차가 일정 이하일 때 해당 동작 모델과 유사한 동작을 행동한 것으로 하였다. 모델에 대한 매칭 확률을 계산하는 방법은 모델링 시에 사용하였던 기호 시퀀스와 상태 시퀀스의 관계성을 이용하여 최적의 상태 시퀀스를 생성하고, 이러한 상태 시퀀스를 이용하여 전체 길이  $T$ 에 대하여 상태 전이 확률과 기호 시퀀스의 발생 확률을 곱하여 구할 수 있다.

인식하고자 하는 동작의 기호 시퀀스  $V^T$ 를 통하여 정규 모델  $\theta$ 에서의 상태 전이 경로(Path)를 구하고,  $t \rightarrow T$ 의 경로 값(전이 확률 값)의  $\text{Sum}(X_1 = A_{\text{Path}})$ 과 모델링 시에 사용한 관측 시퀀스로 구할 수 있는  $\text{Sum}(X_2 = A_{\text{Path}})$  값의 차이로 가장 가까운 모델 접근

```

c ← number of states
begin Path Searching
: Initialize  $t \leftarrow 0, \text{path} \leftarrow \{\}$ 
for t below T
    j ← j + 1
    for j below c
         $\alpha_j(t) \leftarrow b_{jk} v_j(t) \sum_{i=1}^c \alpha_i(t-1) a_{ij}$ 
        j ← j + 1
    endfor
     $j' \leftarrow \arg \langle j \rangle \max \alpha_j(t)$ 
    append  $w_{j'}$  to Path
    t ← t + 1
endfor
return Path
end Path Searching
    
```



이 가능하다. 다음은 최적 경로를 구하는 과정에 대한 알고리즘이다. 최적의 경로는  $T$ 까지의 경로 값의 합이 가장 큰 경로이므로, 경로 검색은 지역탐색이 아닌 전역탐색을 사용해야 한다.

#### 4. 정규 동작과 사용자 동작 간 유사도 측정

유사도는 두 동작에 대한 단계 별 일치 정도를 나타낼 수 있는 척도인데 은닉 마르코프 모델과 같은 확률적 접근은 비교 대상이 되는 두 동작에 대하여 매칭 확률을 나타낼 수는 있지만 전체 시간  $T$ 에 대한 매칭률을 계산하기 때문에 부분적인 단계별 동작 일치에 대한 유사도를 나타내지는 못한다. 본 논문에서는 이러한 부분 동작에 대한 유사도 측정을 위하여 모델링에 사용하였던 기호 시퀀스 재구성하여 부분 일치도 측정을 위한 편집 거리(Edit Distance) 방법 [10-12]에 이용하였다.

기호 시퀀스는 단위 시간당 운동 벡터로 표현되는 방향 정보 형식인데 이를 시간 축에 독립적으로 변환하기 위하여 동일한 기호가 반복되는 경우 이들을 하나로 줄여준다. 이렇게 반복성이 제거된 기호 시퀀스의 형태는 시간 간격에 대한 정보가 제거됨과 동시에 동작의 형태를 나타내는 데이터 형식을 가지게 된다.

본 논문에서는 유사도의 측정 기준으로 두 행동의 지표간의 부분 일치율을 사용하였다. 유사도 계산 방식은 학습에 사용되었던 기호 시퀀스  $S_1$ (모범 행동)와 사용자의 행동에 대한 기호 시퀀스  $S_2$ (사용자 행동)의 점수 행렬을 이용하여 두 시퀀스를 일치 시키는데 발생하는 비용(Cost)의 합을 완전 일치 했을 시에 값  $C(\text{score})$ 에서 감산하는 방식이다. 여기에서  $S_1$ 과  $S_2$ 에 포함되는 시퀀스 집합을  $A$ 라 하고  $A$ 에 공백이 추가된 집합을  $A'$ 라 할 때,  $A'$ 에 있는 임의의 원소  $x$ 와  $y$ 를 정렬할 때 발생하는 비용을  $s(x, y)$ 로 나타낸다. 이 때 두 시퀀스를 일치 시키는데 발생하는 전체 비용은 다음과 같이 표현할 수 있다.

$$A_{\text{cost}} = \sum_{i=1}^l s(S_1'(i), S_2'(i)), l = \text{len}(A) \quad \text{식 (11)}$$

그림 2는  $S_1 = \{6, 2, 7, 6, 3, 6\}$ ,  $S_2 = \{6, 2, 3, 2, 3, 7\}$ ,  $A = \{2, 3, 6, 7\}$ 일 때,  $A'$ 에 속하는 원소들 간의 점수행렬과 정렬  $A$ 의 값을 보여준다.

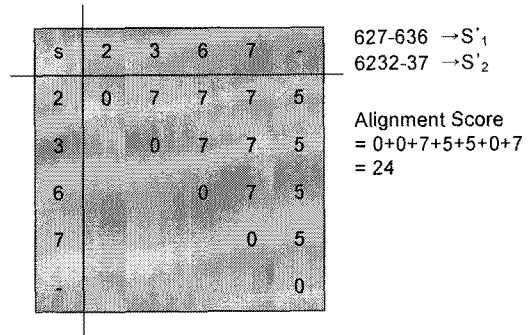


그림 2.  $A'$ 의 점수행렬(Score Matrix) 및 정렬 A 값

점수행렬에서는  $x, y$ 가 동일하면  $s(x,y)$ 는 0이고 동일하지 않으면  $s(x, y)$ 는 0보다 큰 값을 가지게 한다. 이러한 점수행렬을 가지고 가능한 가장 작은 값을 가지는 최적 정렬  $A_0$ 를 구하여 계산한  $A_0$ 의 값을 두 시퀀스 간의 불일치 값으로 사용함으로써 동작 유사도 값  $C(\text{score})$ 를 구할 수 있다. 최적 정렬  $A_0$ 를 구하는 방법과 발생하는  $A_0$ 값은 'Levenshtein distance' 알고리즘[16]을 사용하여 구하였다. 'Levenshtein distance' 알고리즘에서 정의되는 삽입비용과, 삭제비용, 교체비용은 각각 삽입비용(insCost)은  $S_1(i)$ 가 공백이 아니고  $S_2(i)$ 가 공백일 때 발생하는 비용이고, 삭제비용(delCost)은  $S_1(i)$ 가 공백이고,  $S_2(i)$ 가 공백이 아닐 때 발생하는 비용이다. 교체비용(ChangeCost)은  $S_1(i)$ 와  $S_2(i)$ 가 모두 공백이 아니고, 같지 않을 때 발생하는 비용이다. 본 논문에서는 삭제비용과 삽입비용은 5, 교체비용은 7로 설정하였다.

### 5. 실험 결과

#### 5.1 실험 환경

실험에 사용한 PC는 영상 처리, 동작 인식 등 연산 처리를 위한 Host PC 1대와 사용자 영상 출력을 위한 UMPC 1대로 구성하였다. Host PC는 Core2Duo 2.13Ghz, 2GB 메모리 정도의 사양이며, UMPC는 동작 인식 등에 계산은 수행하지 않고 영상 출력에 관련된 연산과 네트워크 연산만을 처리한다. 실험 방법은 3개의 보기 동작에 대하여 각2, 3m 정도 거리에서 10회, 20회, 30회 반복 실험에 의한 동작 인식율을 산출하고, 두 번째로 한 동작을 2~5구간으로 구분하여 각 단계에 대한 부분 동작 실험으로 유사도 점수

계산 결과를 확인하였다.

### 5.2 동작 인식을

실험을 위한 보기 동작은 다음과 같으며, 보기 동작은 동작의 상하, 좌우 이동, 팔·다리 등의 움직임의 기준으로 선택하고 인식 거리에 따른 인식 오인율을 확인하기 위하여 2m, 3m 거리 차를 두고 동작 인식을 실험을 수행하였다.

실험 결과 카메라 위치를 기준으로 정면 운동에 대해서는 95% 이상의 인식율을 보였으며, 거리에 따른 인식율 차이는 크지 않았다. 보기 동작 C에 대해서는 앞으로 이동할 시에 사용자의 행동에서 한걸음 이동 시 마다 좌 이동 후에 우로 무게 중심이 이동되면서 정규 동작 보다 무게 중심이 심하게 이동하면서 이동한 경우에서 인식에 오차가 발생하였다.

### 5.3 동작 유사도 측정

동작 유사도 측정 실험은 보기 동작 A에 대하여 5 단계로 동작을 구분하고 전체 동작을 수행하지 않고 일부 단계 동작만을 행동하였을 때 동작 유사도 결과를 실험해 보았다.

실험 결과 모든 부분 단계들을 모두 수행하였을 경우 거리에 관계없이 정규 동작과 사용자 동작이 정확하게 일치했음을 보였고, 부분 적 단계를 수행하였을 경우 전체 동작에 비례한 점수 결과를 보였다. 동작의 처음 시작을 다르게 하였을 경우(a~c,

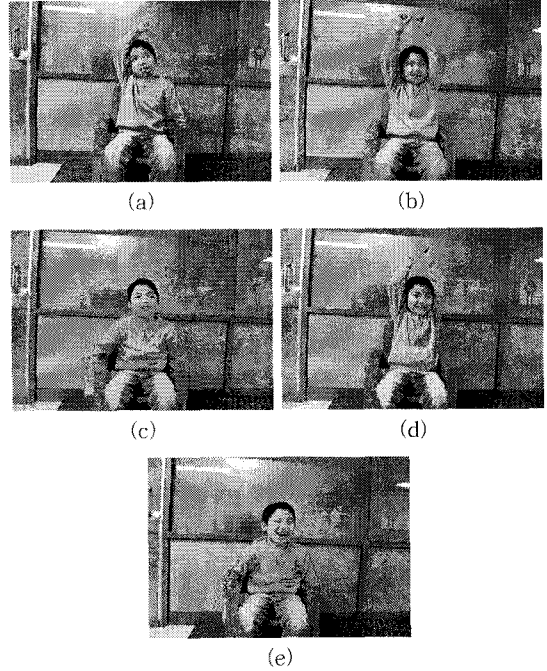


그림 3. 단계별 구분 동작 (a)왼손 들어올리기 (b)오른손 들어올리기 (c)양손 내리기 (d)양손 들어올리기 (e)양손 내리기

표 3. 동작 유사도 측정 결과 점수

구분	2m	3m
a~e	100	100
a~d	72	70
a~c	51	51
d~e	55	52
d	10	0

d~e)에 대해서도 부분적인 동작 유사도 측정이 가능하였다.

본 실험에 사용된 시스템 환경 및 프로그램은 현재 삼육 재활 학교의 지체 장애인 재활훈련에 활용 중이며 실험에 사용된 보기 동작 또한 재활 훈련 시 나리구에 포함된 일부 동작을 사용하였다.

### 5.4 실적용 실험 결과

실제 보행 장애가 있는 아동을 대상으로 보기 동작 C를 수행했고 수행했을 때 인식율 과 유사도에 대한 데이터를 다음과 같이 표 4로 정리 하였다. 동작 자체에 대한 인식률은 실험 때와 비슷한 결과를 보였

표 1. 동작 인식을 실험 동작

구분	동 작	특 징
A	의자에 앉아 양손을 번갈아 들어올리기	팔의 움직임, 몸을 중심으로 팔의 반원 움직임
B	의자에서 일어나기	상하 움직임
C	좌로 3걸음 이동하기	다리의 움직임, 좌우 움직임

표 2. 동작 인식을 실험 결과

구분	10회		20회		30회	
	2m	3m	2m	3m	2m	3m
A	100%	100%	95%	95%	95%	95%
B	95%	95%	95%	90%	90%	90%
C	90%	90%	90%	90%	93%	93%

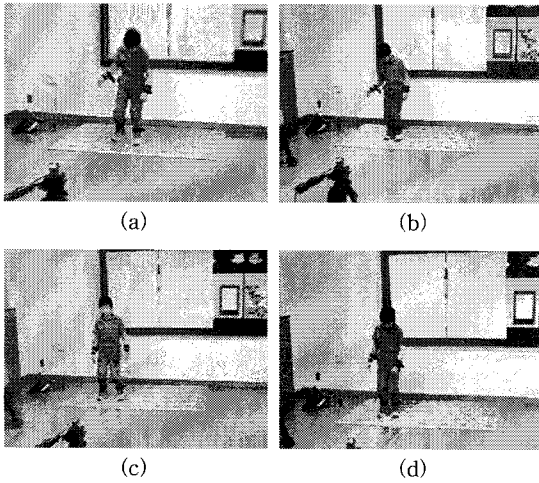


그림 4. 실제 적용 모습

표 4. 실제 적용 결과

	인식률	유사도(점수)
반복	3m	
5	93%	60
10	92%	63
15	92%	65

으며, 유사도의 경우 짧은 이동거리와 보행 장애에 따른 몸 중심 이동의 불안정한 동작으로 인해 비교적 낮은 점수 결과를 보였다.

## 6. 결 론

본 논문은 2차원 영상 특징 기반의 인체 동작 인식 환경과 인식 동작 패턴을 판단하고 측정 지표로서의 유사도 점수를 계산하는 방법을 제시하였다. 제안한 동작 인식 처리 방법은 평균 93%의 높은 인식율을 보였으며, 동작 인식에 그치지 않고 유사도 점수를 계산함으로써 기반이 되는 모범 동작과의 비교 측정이 가능하다. 본 연구의 결과는 동작 인식 기반 게임, 동작의 반복 훈련 및 훈련 달성도 측정, 동작 인식을 이용한 자세 인식 등에 활용 가능하다. 향후 운동 방향에 대한 근사율을 적용하여 방향 구분 시에 구분 경계에 걸치는 운동 방향에 대해서도 인식이 가능하도록 하고 동작에 대한 무게 중심 이동 등을 고려하여 동작 인식율을 높이는 방법에 대하여 연구하고자 한다.

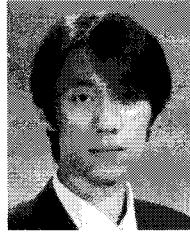
## 참 고 문 헌

- [1] 정순기, 설창환, 원광연, “동작포착을 이용한 인체 동작의 생성,” *한국컴퓨터그래픽스학회 논문지*, 4권, 제 1호, 1998.
- [2] 양문모, “영상처리를 위한 수화인식,” *한국과학재단*, 1997.
- [3] Hiroshi Murase, Shree K and Nayar, “Visual Learning and Recognition 3D object from appearance,” *international journal of computer vision*, Vol.14, 1995.
- [4] Andrew D.Wilson and Aaron F.Bobick, “Recognition and Interpretation of Parametric Gesture,” *Technical Report, MIT*, No.421.
- [5] Thad Eugene Starner, “Visual Recognition of American Sign Language Using Hidden Markov Models,” *Technical Report, MIT*, June, 1995.
- [6] arthur P.Dempster, Nan M.Laird, and Donald B.Rubin, “Maximum-likelihood from incomplete data via the EM algorithm,” *journal of the Royal Statistical Society, Series B*, 1977.
- [7] Geoffrey J.McLachlan and Thiriyambkam Krishnan, “The EM algorithm and Extensions,” Wiley, New York, 1996.
- [8] Michael I. Jordan and Robert A. Jacobs, “Hierarchical mixtures of experts and the EM algorithm,” *Neural computation*, 1994.
- [9] S. B. Needleman and C. D. Wunsch, “A general method applicable to the search for similarities in the amino acid sequence of two proteins,” *Jrnl Molec. Biol*, 1970.
- [10] E. Ukkonen “On approximate string matching,” *Proc. Int. Conf. on Foundations of Comp, Theory*, Springer-Verlag, LNCS 158, pp. 487-495, 1983.
- [11] Stefan Erickeler, Stefan Muller and Gerhard Rigoll, “Improved Face Recognition Using Pseudo 2D Hidden Markov Models,” *Proc. Workshop on Advances in Facial Image Analysis and Recognition Technology*, 1998.
- [12] V. I. Levenshtein, “Binary codes capable of



correcting deletions, insertions and reversals,”  
Doklady Akademii Nauk SSSR 163, pp. 845-848, 1965.

- [13] James William Davis, “Appearance-Based Motion Recognition of Human Actions,” Technical Report, MIT, July 1996.
- [14] Takahiro Watanabe and Masahiko Yachida, “Real timer recognition of Gesture, Gesture Degree Information Using Multi Input Image Sequences,” ICPR, Vol.2, pp. 185~188, 1998.
- [15] Yuanixin Zhu, Yuhuang, Guangyou Xu, Haibing Ren and Zhen Wen, “Vision-based Interpretation of Hand Gestures Bye Modeling Apperance Changes in Mage Sequences,” MVA’98. IAPR Workshop on machine Vision Applications, 1998.
- [16] 김선철, 박인찬, “사용자 제스처 인식을 활용한 유비쿼터스 홈네트워크 인터페이스 체계에 대한 연구,” *감성과학*, 제8권, 제3호, pp. 265-276, 2005.
- [17] 김정현, 노용완, 홍광석, “WPS와 장갑 장치 기반의 동적 제스처 인식기의 구현,” *정보처리학회논문지 B*, 제13-B권, 제5호.
- [18] I. Cohen, N. Sebe, A. Garg, M. S. Lew and T. S. Huang, “Facial Expression Recognition from Video Sequences,” *Proceedings of IEEE International Conference on Multimedia and Exposition*, Vol.2, pp. 121-124, 2002.
- [19] 이철우, 이용재, “파라메트릭 제스처 공간에서 포즈의 외관 정보를 이용한 제스처 인식과 동작 평가,” *멀티미디어학회 논문지*, 제7권, 제8호, pp. 1035-1045, 2004.
- [20] Tae-young Kim, Jun Park and Cheol-su Lim, “Motion Evaluation for Rehabilitation Training of the Disabled,” *IEICE Transactions*, Vol. E91-A, No.9, 2008.
- [21] 김태형, 변혜란, “동영상에서 MGH을 이용한 실시간 다수 동작 인식,” *한국컴퓨터종합학술대회 논문지*, 제33권, 제1(B)호, 2006.



**김 종 호**

2008년 2월 서경대학교 컴퓨터 공학과 학사 졸업  
2008년 2월~현재 서경대학교 컴퓨터 공학과 석사 과정  
관심분야 : 실시간 렌더링, 모바일 3D, 혼합 현실



**윤 요 섭**

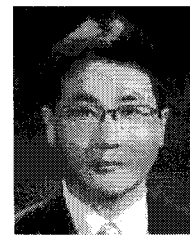
2009년 2월 서경대학교 컴퓨터 공학과 학사 졸업  
2009년 2월~현재 서경대학교 컴퓨터 공학과 석사 과정  
관심분야 : 실시간 렌더링



**김 태 영**

1991년 2월 이화여자대학교 전자계산학과 학사  
1993년 2월 이화여자대학교 전자계산학과 석사  
1993년 3월~2002년 2월 한국통신 멀티미디어연구소 선임연구원

2001년 8월 서울대학교 컴퓨터공학부 박사  
2002년 3월~현재 서경대학교 컴퓨터 공학과 조교수  
관심분야 : 실시간 렌더링, 모바일3D, 증강현실, 블룸그래픽스



**임 철 수**

1985년 서울대학교 계산통계학과 학사 졸업  
1988년 (미) 인디애나주립대 컴퓨터과학과 석사 졸업  
1995년 서강대학교 전산과 박사 졸업  
1997년 3월~현재 서경대학교 컴퓨터 공학과 부교수

2009년 2월~현재 지식경제부 PM  
관심분야 : 차세대컴퓨팅, 멀티미디어 시스템