

논문 2015-52-9-10

## 지화 인식을 위한 계층적 은닉 마코프 모델

( Hierarchical Hidden Markov Model for Finger Language Recognition )

권재홍\*, 김태용\*\*

( Jae-Hong Kwon and Tae-Yong Kim<sup>©</sup> )

## 요약

지화(finger language)는 수화(sign language)에 포함되며, 손의 제스처로 한글의 모음, 자음을 표현하는 언어 체계이다. 한글 지화는 총 31 제스처로 구성되어 있으며, 정확한 인식을 위해서는 하나의 제스처에 대해 학습 모델이 많이 필요로 하게 된다. 대량의 학습 모델이 존재할 경우, 입력 데이터는 많은 공간을 탐색하는데 시간을 소비하게 된다. 따라서 실시간 인식 시스템은 이러한 탐색 공간을 줄이는 것이 가장 중요한 문제로 인식되고 있다. 본 논문에서는 이러한 문제를 해결하기 위해 인식을 저하 없이 탐색 공간을 효율적으로 줄이는 계층적 HMM 구조를 제안하였다. 지화는 손목의 방향성에 따라 총 3개의 범주로 설정, 입력 데이터는 이 범주 안에서 모델을 검색하게 된다. 이러한 사전 분류를 진행하여 비슷한 한글 지화의 분별력을 확립하게 되며 탐색 공간 또한 효율적으로 관리되므로 실시간 인식 시스템에 적용 가능하다. 실험 결과, 제안된 방법은 일반적인 HMM 인식 방법보다 평균 3배 정도의 시간을 단축할 수 있었고, 비슷한 한글 지화 제스처에 대해 오인식 또한 감소하였다.

## Abstract

The finger language is the part of the sign language, which is a language system that expresses vowels and consonants with hand gestures. Korean finger language has 31 gestures and each of them needs a lot of learning models for accurate recognition. If there exist mass learning models, it spends a lot of time to search. So a real-time awareness system concentrates on how to reduce search spaces. For solving these problems, this paper suggest a hierarchy HMM structure that reduces the exploration space effectively without decreasing recognition rate. The Korean finger language is divided into 3 categories according to the direction of a wrist, and a model can be searched within these categories. Pre-classification can discern a similar finger Korean language. And it makes a search space to be managed effectively. Therefore the proposed method can be applied on the real-time recognition system. Experimental results demonstrate that the proposed method can reduce the time about three times than general HMM recognition method.

**Keywords :** 손 인식(hand recognition), 립모션 (leap motion), 지화 (finger language), 은닉 마코프 모델 (hidden markov model), 계층 모델(hierarchical model)

## I. 서론

인간-컴퓨터 상호작용 (Human-Computer

\* 정희원, \*\* 평생회원, 중앙대학교 첨단영상대학원 영상학과

(Graduate School of Advanced Image Science, Multimedia, and Film, Chung-Ang University)

© Corresponding Author(E-mail: kimty@cau.ac.kr)

※ 이 논문은 2014년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(2013-009166, NRF-2014S1A5B6037633).

Received ; May 12, 2015 Revised ; August 12, 2015

Accepted ; September 3, 2015

Interaction)에 대한 연구가 국·내외에서 활발히 진행되면서 컴퓨터는 단순한 연산 결과만을 출력하는 것이 아닌, 인간과의 소통을 중심으로 기술은 진화되고 있다. 그중에서도 제스처 인식 기술은 인터랙션 아트, 게임, 가상 현실, 사용자 인터페이스(User Interface)의 개선 등 활용 범위가 매우 넓고 이에 대한 연구 중요성이 증가하고 있다. 이러한 연구가 진행되는 가운데 최근 인간과 컴퓨터 간의 의사소통이 보다 자유롭고 지적인 시스템을 설계하기 위해서는 수화, 지화와 같은 시각언어의 인식 또한 중요한 문제로 대두되고 있다. 하지만 복잡한 환경에서의 사람 손의 추출 및 제스처

인식을 정확히 한다는 일은 아주 어려운 일이며, 이러한 손 제스처를 인식하기 위한 기존의 연구들은 영상에서의 손의 2차원 특징을 추출하여 패턴을 분석하는 방법<sup>[1]</sup>, Data-Glove를 사용하여 손의 형상을 분석하는 방법<sup>[2]</sup>과, 손의 3D 모델을 사용하거나 스테레오 영상을 획득하여 전처리 과정을 통해 추출된 손의 3차원 기하학적인 정보를 분석하는 방법<sup>[3]</sup>으로 분류할 수 있다. 하지만 이 세 가지 방법에는 각각의 문제점들을 가진다. 영상 기반으로 손 제스처 인식을 하는 경우 입력 영상 내에서의 피부색 제약, 단일 배경, 조명 변화 등 실험 환경에 너무 의존적이고, 정확한 인식을 위한 전처리 단계에서 많은 연산이 필요하게 되므로 실시간 시스템에 적용하기 어렵다. Data-Glove를 이용할 경우 사용자는 고가의 장비를 갖추어야 하며, 또한 장갑을 착용해야 하는 불편함이 존재하므로 HCI 인터페이스 구축에 많은 제약이 따른다. 마지막으로 손의 3차원 기하 정보를 이용하는 방법은 스테레오 영상을 이용하여 영상 데이터의 정합, 복잡한 손의 3차원 운동 모델을 만들기 위한 많은 연산량이 발생하므로 고성능의 하드웨어, 또는 많은 시간이 요구되는 문제점이 있다. 이 밖에도 손 제스처를 인식하기 위해서는 제스처의 스캐닝, 제스처의 반복성과 결합성 등 여러 가지 문제를 해결해야만 한다. 지화·수화 실시간 언어 인식 시스템의 가장 큰 문제점으로는 실시간 인식을 위하여 다량의 모델 검색을 효율적으로 관리할 수 있어야 한다. 한글 지화에는 비슷한 제스처가 존재하고, 이에 오인식 하는 문제점이 있다. 본 논문에서 제안하는 방법은 사전 분류를 실시하여 DB검색 시간 감소, 비슷한 제스처 모델을 검색에서 제외하여 인식률이 향상되는 방법을 제시한다. 손을 추적하는 소형 모션 컨트롤러(Leap Motion)를 사용하여 특별한 전처리와 장비의 착용 없이 3차원 상에서 손에 대한 데이터를 획득, 손에 대한 데이터는 간단한 전처리를 통해 특징점으로 추출되고 HMM(Hidden Markov Mode)으로 모델링 된다. 이후 새로운 입력 데이터는 손목의 방향성에 따른 사전 분류를 실시, 분류된 범주에서 DB모델과 매칭을 실시하여 최종적으로 인식을 한다.

본 논문의 구성은 다음과 같다. II장에서는 한글 지화, Hiden Markov Model 및 손 데이터 획득 방법에 대해 설명하고, III장에서는 제안된 시스템의 요소들을 자세히 기술하고, IV장에서는 제안된 제스처 인식 시스

템과 일반적인 제스처 인식 시스템에 대한 실험 결과 분석을 논하며, 마지막 V장은 이후 연구와 결론을 맺는다.

## II. 지화 인식과 HMM

### 1. 지화 인식 시스템

지화 언어는 청각장애인들이 손동작으로 나타내는 시각적 언어이며 수화에 속하며, 한글 지화 문자는 한글의 자음, 모음을 표현하는 음소 단위로 글자를 표현할 수 있다. 한글 지화는 총 31 가지의 제스처로 구성되어 있으며, 각각의 제스처에 대한 문자는 그림 1과 같이 표현된다.

수화·지화 인식 시스템에 대한 연구는 세계 각국에서 다양한 연구들이 진행되고 있다. 미국은 수화를 인공 신경망에 적용한 수화 인식 시스템<sup>[4]</sup>에서 14개의 제스처에 대해 평균적으로 84%의 인식률을 달성하였다고 한다. 일본의 경우, Data-Glove를 사용해 46개의 지문자 중 34개를 인식하는 시스템, 영상에서 특징을 추출하여 16개의 일본 수화에 대하여 평균적으로 82% 인식률을 보이는 시스템<sup>[5]</sup>에 대한 연구가 진행되었다. 한국에서는 Data-Glove를 이용한 연구<sup>[6]</sup>와 영상에서 수화·지화를 인식하는 방법<sup>[7~8]</sup> 등, 세계에서 이에 대한 연구가 활발히 이루어지고 있다.

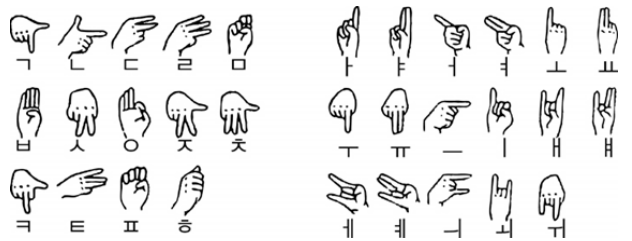


그림 1. 한글 지화 (31 제스처)  
Fig. 1. Korean finger language (31 gestures).

### 2. 손 데이터 획득 방법

HCI 시스템 설계 및 구성을 하기 위해서는 화상 카메라, Data-Glove, Kinect 등, 다양한 장비들을 이용하여 구성 가능하다. 본 논문에서는 Leap Motion 장비를 이용하여 실험을 진행하였다. Leap Motion은 기본적으로 제스처 기반의 유저 인터페이스를 제공하는 장치로서 Microsoft 사의 Kinect 보다 약 200배 정도 높은 감도를 가지며, 0.01mm의 정밀도로 손목과 손가락, 기타

물체를 인식할 수 있는 입력장치이다<sup>[9]</sup>. 3개의 적외선 LED를 사용하여 적외선 반사 정보, 2개의 적외선 카메라라는 깊이 정보를 정합하여 오브젝트의 빠르고 정밀한 데이터를 제공한다. 손의 인식 가능 범위는 장치 위에서 대략 60cm 정도로 제한되며, 매 프레임마다 손목에 대한 정보, 손가락에 대한 정보가 데이터로 제공된다. 개발사는 3D 디스플레이어에서의 객체 제어, 멀티 디스플레이 제어, 가상 현실 제어 등 향후 미래 지향적인 인터페이스를 구성할 수 있는 입력 장치라고 설명한다. 화상 카메라, Kinect, Data-Glove 장비들을 이용하여 손에 대한 정보를 획득하고 가공하기 위해서는 강력한 알고리즘과 그에 따른 많은 연산량이 요구되지만, Leap Motion을 사용하면 손에 대한 정보, 추적 상태 등 전 처리를 따로 진행하지 않고 빠르고 정확하게 정보를 획득할 수 있는 장점이 있다.

### 3. Hidden Markov Model (HMM)

은닉 마코프 모델(HMM)이란 마코프 모델에서의 상태가 은닉됐다는 의미이다. 마코프 모델의 모델링 능력에 한계가 있어 보다 복잡한 현상이나 과정은 제대로 모델링 되지 못하였다. 이후 지속적인 연구를 통해 은닉 마코프 모델(HMM)로 발전하게 되었다. HMM은 이중 확률 모형의 하나로서 입력된 신호의 패턴을 인식하거나 필기체, 음성, 제스처 인식 등 다양한 문제 해결에 널리 사용된다. 객체 패턴을 나타내는 관찰 심볼은 스트링 형태로 분석하여 이를 통계적 모형으로 학습 시킨 후, 입력된 신호에 대해 관측 확률이 얼마인가를 계산하는 방법이다. 모델의 학습은 관측 열에 대하여 최대 확률을 가지게 되는 통계적 결과가 생성된다. HMM은  $\lambda = (A, B, \pi)$ 으로 표현하며 3가지 매개변수를 가진다.

#### 가. 상태 전이 확률, 관측 확률, 초기 상태 확률

상태 전이 확률은 HMM이 다음 상태를 결정하는 확률이다. 시간 t에서  $s_i$  상태에서 t+1에  $s_j$ 로 이동할 확률을  $a_{ij}$ 라고 하며 HMM의 모든 상태 수(n)에 대하여 합은 1이며  $(\sum_{j=1}^n a_{ij}=1)$ , 식 (1)로 표현할 수 있다.

$$a_{ij} = P(q_{t+1} = s_j | q_t = s_i) \quad (1)$$

여기에서  $i, j$ 는  $1 \leq i, j \leq n$  범위를 갖는다. 관측 확률은 HMM의 어느 상태에서 기호가 관측될 확률이

다.  $b_j(v_k)$ 는 상태  $s_i$ 에서 기호  $v_k$ 가 관측될 확률이며, m은 기호의 개수이다. 상태  $s_i$ 에서의 모든 관측확률의 합은 1이며  $(\sum_{k=1}^m b_j(v_k)=1)$ , 식 (2)로 표현 할 수 있다.

$$b_j(v_k) = P(o_t = v_k | q_t = s_j) \quad (2)$$

이때  $j, k$ 는  $1 \leq j \leq n, 1 \leq k \leq m$ 의 범위를 갖는다. 초기 상태 확률은 HMM이 초기에 어느 상태에서 시작하는지 결정되는 확률이다.  $\pi_i$ 는 상태  $s_i$ 에서 시작하는 확률이며 모든 상태 수(n)의 합은 1이며  $(\sum_{i=1}^n \pi_i = 1)$ , 식 (3)으로 표현할 수 있다.

$$\pi_i = P(q_1 = s_i) \quad (3)$$

이때  $i$ 는  $1 \leq i \leq n$ 의 범위를 갖는다.

하지만 HMM을 실제 시스템에 적용하기 위해서는 세 가지 (확률 평가, 디코딩, 학습) 문제점을 해결해야 한다. 세 가지 문제에 대한 해결 방법은 다음과 같다.

#### 나. HMM의 세 가지 문제(확률 평가, 디코딩, 학습)

확률 평가 문제는  $P(O|\lambda)$ 를 계산하는 문제이다. 관측 벡터  $O = \{o_1, o_2, o_3, \dots, o_t\}$ 와 은닉 마코프 모델  $\lambda = (A, B, \pi)$ 가 주어질 경우 이를 평가해야 된다. 단순히 확률론적인 계산을 하게 되면 모든 상태열에 대해서 관측 확률을 구해야 하기 때문에 엄청난 연산량을 요구한다. 하지만 이러한 문제점은 forward-backward 알고리즘<sup>[10]</sup>을 통해 연산량을 효율적으로 관리할 수 있다. 디코딩 문제는 모델  $\lambda$ 에서 관측 벡터  $O$ 에 해당하는 최적의 상태열  $Q = \{q_1, q_2, q_3, \dots, q_t\}$ 를 찾는 것이다. 이 문제 또한 최적의 상태열을 계산하는 과정에서 엄청난 연산량을 요구한다. 동적 프로그래밍 방법인 Viterbi 알고리즘<sup>[10]</sup>을 이용하여 이러한 연산량을 줄일 수 있었다. 학습 문제는 관측 벡터  $O$ 의 확률을 최대화하는 모델  $\lambda$ 을 구하는 문제이다. 앞서 제시된 두 문제의 경우 연산량에 의한 문제이지만, 학습은 전혀 다른 문제이다. 본 논문에서는 이 문제를 해결하기 위해 Baum-Welch<sup>[10]</sup> 알고리즘을 사용하였다. Baum-Welch는 EM(Expectation Maximization)알고리즘을 이용하여 학습한다. 관측열은 두 개의 은닉 변수를 사용하여 기댓값을 구하는 E 단계와, 우도를 최대화하는 M 단계의

반복적인 연산을 통해 모델을 추정한다.

### III. 계층적 지화 인식 시스템

본 장에서는 사전 분류를 적용한 한글 지화 인식 시스템을 기술한다. 전체적인 시스템 개요도는 그림 2와 같다. 먼저 입력된 데이터는 손목의 방향성 정보와 각 손가락별 위치 정보를 가지며, 손가락 위치정보는 간단한 전처리 과정을 통해 체인코드로 변환하며, 손목의 방향성 정보를 이용하여 한글 지화 모델을 사전 분류를 실시한다. 획득한 체인코드는 사전 분류가 실시된 후, 카테고리 내에서의 DB의 모델들과 매칭을 실시하여 최종적으로 인식한다.

#### 1. 한글 지화 인식

한글은 기본 자음, 복합 자음, 모음, 복합 모음을 합하여 총 40가지의 음소로 구성되며 이를 조합하여 많은 단어를 표현할 수 있다. 하지만 실시간 인식 시스템에서 많은 제스처를 정확히 인식하는 일은 시스템의 자원, 시간적 측면에서 비효율적인 일이다. 한글의 기본 자음·모음으로 모바일 기기에서 채팅 및 검색(천지인 키보드)을 할 수 있듯이, 본 논문에서 인식하고자 하는 지화들은 총 31가지 제스처 이지만 정확한 인식과 시스템 자원의 효율적 관리를 위하여 한글의 기본 자음·모음을 천지인 키보드에 적용하여 총 11개의 지화 제스처에 대하여 인식한다. 그림 3의 (b)는 모음·자음 9개와 칸 띄우기, 획 추가 입력은 지화 제스처에서 임의로 정의하여 총 11개의 구성되며, (b)의 모음·자음 제스처 요소들은 한국 지화에서 참조하였다.

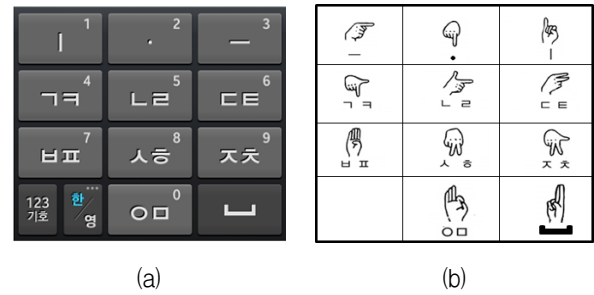


그림 3. (a) 천지인 키보드, (b) 천지인 키보드에 적용한 한글 지화.  
Fig. 3. (a) Tenchijin keyboard, (2) Tenchijin applied to the keyboard of Korean finger language.

#### 2. 특징 추출

##### 가. 손가락 체인코드 추출

최근 연구<sup>[11]</sup>에 따르면 제스처 인식에 있어서 사용되는 특징점은 크게 오브젝트의 위치, 동선, 움직인 거리로 분류할 수 있다. 각각 특징 값을 제스처 인식 시스템에 적용한 결과, 추출된 오브젝트의 위치는(Location) 46%, 오브젝트의 동선은(Angle) 87%, 오브젝트의 움직인 거리는(Velocity) 32%의 인식률을 보인다. 결론적으로 오브젝트가 움직이는 동선의 각도(Angle)가 제스처 인식 시스템의 가장 영향력 있는 특징점이 된다. 또한 한국수화 제스처 인식을 위한 특징 추출 연구를<sup>[12]</sup> 참조하였고, 본 논문에서는 이러한 연구를 바탕으로 손가락 동선을 제스처 인식 시스템에서 특징으로 사용할 것이다. Leap Motion 장비는 기본적으로 손에 대한 다양한 데이터를 제공하지만 본 논문에서는 각 손가락 끝의 위치와 사전 분류의 특징으로 사용되는 손목의 회전력만 데이터로 사용한다. 오브젝트의 동선의 각도를 계산하기 위해서는 먼저 손가락 XYZ 좌표를 얻은 다음, 이전

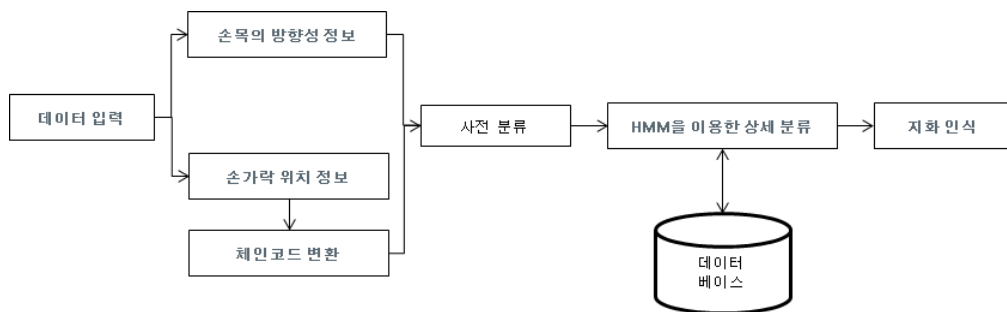


그림 2. 제안된 한글 지화 인식 시스템  
Fig. 2. The proposed Korean finger language recognition system.

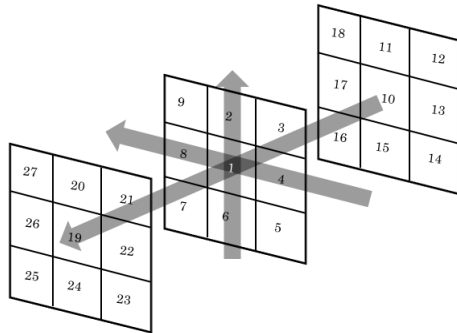


그림 4. 27방향 3차원 체인코드  
Fig. 4. 27-direction of 3 dimensional chain-code.

프레임 번호	엄지	검지	중지	약지	소지
1	11	16	28	28	28
2	6	21	28	28	28
3	1	3	28	28	28
4	26	19	28	28	28
5	8	1	28	28	28
6	8	1	28	28	28
7	1	1	28	28	28
8	1	1	28	28	28
9	1	1	28	28	28
10	1	1	28	28	28
11	1	1	28	28	28
12	1	1	28	28	28
13	3	28	28	28	28
14	3	28	28	28	28
15	1	28	28	28	28
16	24	28	28	28	28
17	16	12	28	28	28
18	17	19	28	28	28

그림 5. 추출된 ‘ㄱ,ㅋ’ 제스처의 체인코드  
Fig. 5. Extracted ‘ㄱ,ㅋ’ chain-code of gesture.

프레임과의 차분을 통하여 3차원 27방향과 손가락이 추적되지 않는 상태 1을 더하여 총 28방향 체인코드를 생성한다.

획득한 체인코드를 아래와 같은 5차원 벡터로 구성하고, 양자화를 진행 후 Baum-Welch 학습 알고리즘을 이용하여 모델링 한다.

$$V = \{\text{엄지, 검지, 중지, 약지, 소지 체인코드}\}$$

앞서 선언한 천지인 키보드에서의 ‘ㄱ,ㅋ’의 제스처는 엄지와 검지의 움직임만 존재한다. 그림 5는 실험에서 추출된 제스처 ‘ㄱ,ㅋ’ 손가락별 프레임에 따른 체인코드를 보여준다. 프레임에 따라서 엄지와 검지의 체인코드는 변화하지만, 나머지 손가락들의 체인코드는 추적되지 않는 상태(28)를 나타내고 있다.

#### 나. 손목의 방향성 추출

Leap Motion에서의 손목 회전력에 관한 데이터는

Pitch, Roll, Yaw 3가지로 구분된다. 제안된 제스처 인식 시스템에서 사용될 데이터는 Pitch 값, 즉 z축 및 y-z 평면상에 투영된 벡터의 사이 각이다. Pitch 값을 이용하여 사전 분류를 하는 방법은 다음 장에서 기술한다.

#### 3. 손목 방향에 의한 사전 분류

본 논문에서는 입력 데이터와 DB모델 간의 매칭을 시행하기 전 사전 분류하는 방법을 제안한다. 제안하는 한글 지화의 카테고리는 손목의 방향성에 의해 그림 6과 같이 나눌 수 있다. 입력 데이터는 손목의 Pitch 값과 DB모델의 Pitch 값으로 탐색할 카테고리를 선정한다. 지화는 손목의 방향성을 3개의 부류로 분류 가능하며 사전 분류를 하는 이유는 다음과 같다.

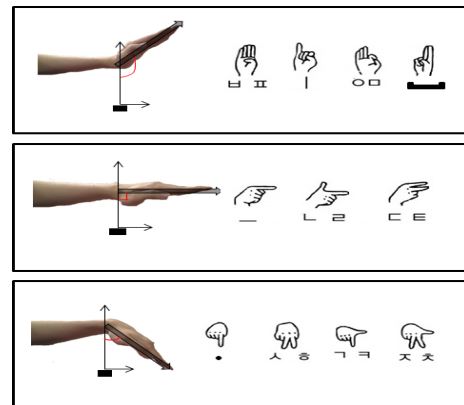


그림 6. 손목의 방향성에 따른 한글 지화 분류  
Fig. 6. Korean finger language classified by the orientation of the palm.

#### - 탐색 공간 축소

일반적인 HMM인식 시스템에서의 입력 데이터는 DB모델을 전체적으로 검색하지만 제안된 방법은 3가지 카테고리 중 손목의 방향성 정보를 이용하여 카테고리 내의 모델만을 검색하여 전체적인 탐색 공간을 축소 시킨다. 이에 대량의 DB 모델이 존재하는 정교한 시스템의 경우 일반적인 방법과 제안된 방법은 탐색 공간의 많은 차이를 가지게 된다.

#### - 인식률 향상

한글 지화의 제스처 중 비슷하여 정확히 식별하기 어려운 형태를 띠고 있다. 예를 들어 ‘ㄷ-ㅌ’ 과 ‘ㅅ-ㅎ’ 등 손목의 방향성을 제외하면 비슷한 제스처로

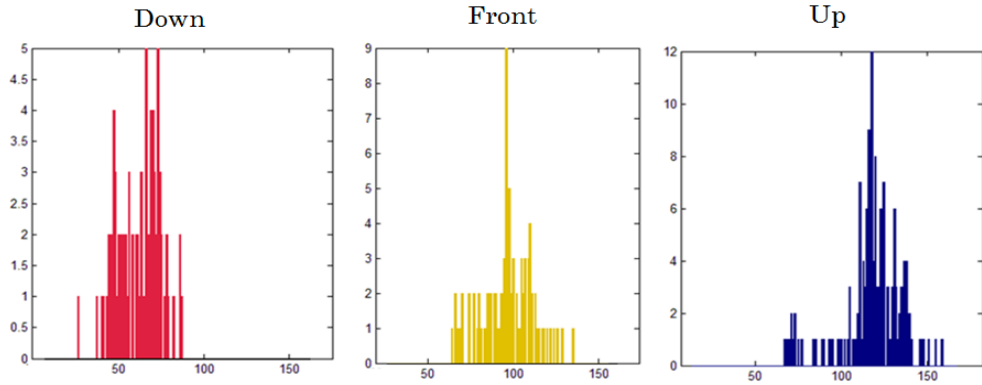


그림 7. 학습 데이터에서의 추출된 Pitch값의 히스토그램  
 Fig. 7. Pitch values extracted from the histogram of the learned data.

볼 수 있다. 이러한 환경은 오인식 확률을 높여주게 되며 이를 해결하기 위해서는 관측 벡터에 새로운 정보를 추가해야 될 것이다. 만약 관측 벡터의 차원이 커지면 커질수록 모델 검색에 대한 연산과 Baum-Welch 학습의 문제점인 과적합 문제점도 함께 해결해야 한다. 이러한 일련의 문제점 없이 모델을 카테고리 나누어 선별적으로 분류하는 방법은 비슷한 제스처에 대해 오인식을 줄여주게 되며 결과적으로 인식률 향상이라는 결과를 가져오게 된다.

사전 분류를 진행하기 위해서는 사전에 학습된 모델의 Pitch값 분포를 획득하며, 그림 7은 각 카테고리 내에서의 각 제스처 대한 Pitch값을 히스토그램으로 표현한다. 각 카테고리의 평균 Pitch값은 67.3, 98.9, 113.2로 나타나며, 이 정보들을 이용하여 각 카테고리의 Pitch값을 정규분포 모델화한다. 이후 새로운 입력 데이터는 이 카테고리의 정규 분포 확률이 높은 곳에서 탐색을 시작하게 된다. 전체적인 사전 분류 진행과정은 그림 8과 같다. 그림 8에서는 데이터가(빨동그라미로 표시한 부분) 입력이 된다면, 3가지 카테고리에 대한 정규분포 확률은 Down에서 가장 높다. 따라서 Down카테고리에서 모델을 탐색하게 되며 세부적인 매칭을 실시한다. 본 논문에서는 11가지 제스처 대상으로 사전 분류하였

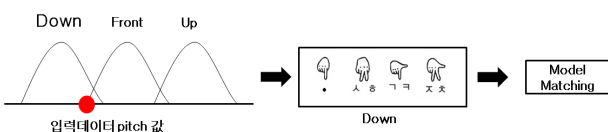


그림 8. 사전 분류의 과정  
 Fig. 8. The pre-classification process.

지만 한글 지화 전체를 대상으로 하는 시스템에서도 충분히 사용 가능하다.

#### 4. HMM 구조 및 초기화 설정

일반적인 HMM의 토폴로지모델은 어느 상태가 모델의 다른 상태로부터 도달할 수 있는 ergodic모델(완전연결)이다. 하지만 ergodic모델은 상태 전이에 대한 많은 경우의 수를 가지므로 데이터를 쉽게 잃을 수 있다. 이전 연구<sup>[14]</sup>에서는 제스처 인식 시스템에서 ergodic모델 보다 Left-Right모델을 사용하는 것이 더 효과적이라고 나타났다. 이전 연구를 바탕으로 본 논문에서도 Left-Right모델을 사용하였으며, 상태의 수는 3~6까지 변화시키며 실험을 진행하였다. 또한 이전에서 획득한 특징 벡터는 HMM에서 관측열로 사용 가능하도록 양자화 처리를 해야 한다. 본 논문에서는 K-means<sup>[13]</sup> 알고리즘을 이용하여, 수집한 특징 벡터들을 K 개의 클러스터로 분류 후 그 중심(Center)을 Codebook에 기록, 새로운 입력 벡터는 Codebook을 참조하여 가장 가까운 거리 있는 클러스터로 분류한다. 즉 Codebook은 새로운 입력 벡터를 관측열로 변환시키기 위해 사용된다. 실험에서 사용되는 K는 64의 값을 가진다. 마지막으로 획득한 제스처의 데이터를 모델링하기 위해서는 HMM의 매개변수들의 초기 값들을 설정하여야 한다. 아래의 매개 변수 초기화는 상태 수 5를 기준으로 기술하였다.

먼저 초기 확률은  $\pi = (1, 0, 0, 0, 0)^T$  같이 설정할 수 있다. 왜냐하면 여러 상태들 중 첫 번째 상태부터 시작하기 위함이다.

Left-Right 모델에서의 상태 전이 행렬 A의 초기 값들은 식 (4)와 같이 설정할 수 있다. Left-Right 모델에

서의 상태 전이는 이전 상태에 대해 전이가 불가하고, 모든 상태 전이에 대하여 동일한 확률의 값으로 설정한다. 앞 절에서 설명하였듯이 각 행에서의 모든 상태 전이 확률의 합은 1이 되어야 하므로  $a_{ij}$ 는  $a_{ij} = \frac{1}{n}$  로 나타낼 수 있다.  $n$ 는 각 행에서의  $a_{ij}$ 의 수이다.

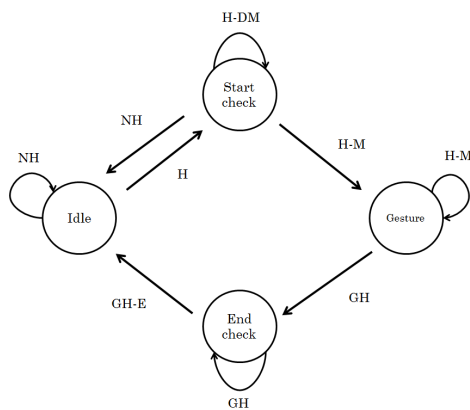
$$A = \begin{pmatrix} a_{ij} & a_{ij} & a_{ij} & a_{ij} & a_{ij} \\ 0 & a_{ij} & a_{ij} & a_{ij} & a_{ij} \\ 0 & 0 & a_{ij} & a_{ij} & a_{ij} \\ 0 & 0 & 0 & a_{ij} & a_{ij} \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad (4)$$

마지막으로 관측 행렬  $B$ 의 모든 요소는 다른 모든 상태에 대해 동일한 값으로 초기화한다. 이는  $B = b_{im}$  와  $b_{im} = \frac{1}{M}$  로 표현할 수 있으며, 여기서  $i, M$ 은 각각 상태의 수, 심볼의 수를 나타낸다.

#### IV. 실험

##### 1. 실험 환경

본 논문에서 사용된 실험 데이터 셋은 11개의 제스처에 대해 각 5개의 학습 데이터와 1100개의 입력 데이터



심볼	손의 추출 상태	손의 움직임
H	있음	있음
NH	없음	없음
H-DM	있음	없음
H-M	있음	있음
GH	있음	없음
GH-E	있음	없음

그림 9. 스팟팅을 위한 심볼 상태  
Fig. 9. Status symbols for spotting.

를 기준으로 진행하였다. 실험은 제안된 방법(H-HMM)과 사전 분류를 진행하지 않는 일반적인 HMM 인식 시스템에서 상태 수를 3~6까지 변경하며 실험을 진행하였다. 또한 실시간에서의 제스처의 시작과 끝을 알 수 없으므로 실험에서는 다음과 같은 조건<sup>[11]</sup>으로 제스처의 스팟팅을 시행하였다. 그림 9에서 볼 수 있듯이 대기 상태에서 손이 움직이면(H-M) 제스처의 시작으로 간주하고 입력 신호를 저장한다. 그리고 손의 제스처가 끝나면 손의 Grab(GH) 상태로 제스처의 끝을 정의한다. GH 상태가 어느 정도의 프레임 동안 유지가 되면 저장하였던 입력신호를 DB 모델과 비교하고, 다시 대기 상태로 돌아가 제스처의 인식을 준비한다.

##### 2. 실험 결과

본 절에서는 제안된 방법(H-HMM)과 일반적인 HMM 인식 방법에 대한 각 카테고리의 인식을 실험, 최적의 시스템을 구성하기 위해 HMM의 상태 수를 변화 시키며 실험한 결과는 그림 10, 11, 12와 같다.

그림 10, 11, 12에서 보여주듯이 인식률에서 일반적인

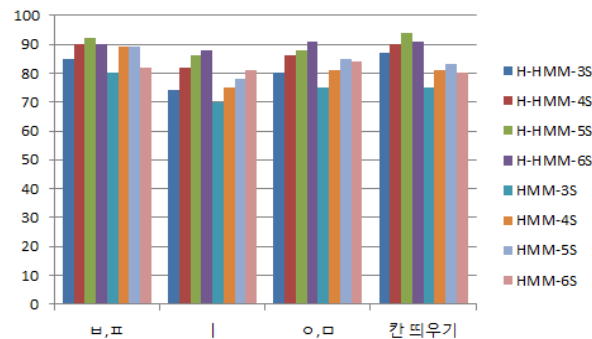


그림 10. Up 카테고리 제스처 인식률  
Fig. 10. 'Up' Categories recognition rate.

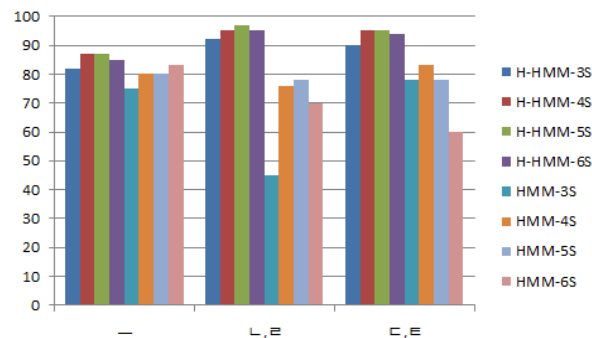


그림 11. Front 카테고리 제스처 인식률  
Fig. 11. 'Front' Categories recognition rate.



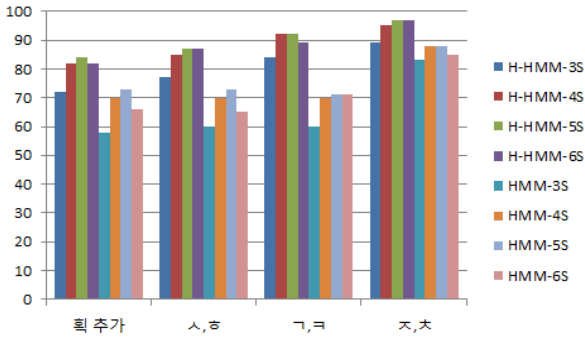


그림 12. Down 카테고리 제스처 인식률  
 Fig. 12. 'Down' Categories recognition rate.

표 1. 각 방법의 상태수에 따른 평균 인식률  
 Table 1. The average recognition rate of each method.

Ratio(%)	평균 인식률
H-HMM-3S	82.90
H-HMM-4S	89
H-HMM-5S	90.81
H-HMM-6S	89.90
HMM-3S	69
HMM-4S	78.45
HMM-5S	79.63
HMM-6S	75.18

표 2. DB 모델 매칭 속도  
 Table 2. DB model Matching speed.

H-HMM-3S	0.0076/s
H-HMM-4S	0.0077/s
H-HMM-5S	0.0078/s
H-HMM-6S	0.0079/s
HMM-3S	0.0192/s
HMM-4S	0.0194/s
HMM-5S	0.0195/s
HMM-6S	0.0197/s

HMM 방법의 인식률이 제안된 방법과 많은 차이를 보인다. 일반적인 HMM 인식 방법에서는 사전 분류를 진행하지 않고 입력신호에 대해 모든 모델과 비교하므로 비슷한 모델을 오인식하는 문제가 발생하였다. 즉 손목의 방향성을 제외한 비슷한 제스처(“ㄴ, ㄹ” 과 “ㄱ, ㅋ”, “ㅡ” 와 “ㅣ” 그리고 “획 추가” 등)에서의 인식률에서는 많은 차이를 보였으며 “ㅇ, ㅁ”, “ㅂ, ㅍ”과 같은 분별성이 있는 제스처에 대해서는 인식률이 적은 차이를 보인다. 또한 HMM를 이용한 한글 지화 제스처 시스템에서 상

태 수를 변화 시키며 실험한 결과, 상태 수 5(5S)에서 두 방법 모두 최상의 인식률을 달성하였다. 이는 HMM 구조에서 상태 수가 너무 적으면 인식률에 대해 저하가 생긴다. 또한 계속적으로 상태 수를 늘려보았으나 미미한 인식률 차이, 매칭 시간이 증가하는 것을 실험 결과에서 보여주고 있다. 각 방법의 평균적인 인식률은 표 1과 같이 나타난다. 표 2는 각 방법에 대해 DB모델 하나에 대한 검색 처리 속도를 나타낸다. 제안된 방법은 사전 분류를 진행하여 카테고리 범주 안에서 DB모델과의 검색을 처리하였기 때문에 일반적인 HMM 인식 방법과 평균적으로 2~3배 정도의 시간이 절약되게 된다. 또한 각 방법의 상태수가 증가함에 따라 DB모델과 매칭 되는 시간이 증가한다. 만약 한글 지화 인식 시스템에서 정확한 인식을 위해 다양하고 많은 DB모델을 대상으로 실험을 하였으면 인식률과 DB모델 검색 속도는 더욱더 많은 차이를 보이게 될 것이다.

#### IV. 결론

본 논문에서 손목의 방향, 각 손가락의 위치 이 두 가지의 기본 정보만을 이용하여 지화의 11가지 제스처에 대한 인식 시스템을 구축하여 실험하였다. 한국 지화를 손목의 방향성에 따라 분류하여, HMM 제스처 인식 시스템에서의 오인식 감소, 효율적인 매칭을 하는 것이 이 논문의 주된 목표이다. 제안한 방법의 우수성을 확인하기 위해 일반적인 HMM 인식 시스템과 비교 실험을 진행하였다. 최적의 상태(5S)에서 제안된 방법의 인식률은 평균 90%로서 일반적인 HMM 시스템 보다 약 10% 정도의 높은 인식률을 달성하였고 실시간 시스템에서 가장 중요한 DB모델 비교 처리 속도는 평균 2~3배 정도 빠른 것으로 나타났다. 실시간 시스템에서 보다 자연스럽게 제스처의 스캐핑을 실시한다면 향후 연구에서 더욱 정확한 인식 시스템을 구축할 것으로 보인다.

#### REFERENCES

[1] Kazuyuki Imagawa, Shan Lu, “Color-Based Hands Tracking System for Sign Language Recognition”, Automatic Face and Gesture Recognition IEEE International Conference, pp. 462-467, 1998.



[2] Ji-Hwan Kim, Tae-Seong Kim, "3-D Hand Motion Tracking and Gesture Recognition Using a Data Glove", IEEE International Symposium on Industrial Electronics, pp. 1013-1018, 2009.

[3] Jakub segen, Senthil Kumar, "Shadow Gestures:3D Hand Pose Estimation Using a Single Camera.", IEEE Computer Society Conference on, Vol. 1, pp. 479-485, 1999.

[4] Manjula, Waldron, Soowon Kim, "Isolated ASL Sign Recognition System for Deaf persons", Rehabilitation Engineering, IEEE Transactions on, Vol. 3, pp. 261-271, 1995.

[5] Yamaguchi, "Japanese sign language recognition system using information infrastructure", International Joint Conference of the Fourth IEEE International Conference on Fuzzy Systems and The Second International Fuzzy Engineering Symposium, Vol. 5, pp. 65-66, 1995.

[6] Seungki Min, Sanghyeok Oh, Gyoryeong Kim, Taehyun Yoon, Chungyu Lim, Yunli Lee, Keechul Jung, "Optimize Data Glove-based System for Korean Finger Spelling Recognition", Korea Computer Congress, Vol. 34, pp. 237-241, 2007.

[7] Yang HeeDeok, Lee Seong-Whan, "Automatic Spotting of Sign and Fingerspelling for Continuous Sign Language Recognition", The Korean Institute of Information Scientists and Engineers, Vol. 38, pp 102-107, 2011.

[8] Min-Ji Kang, Eun-Sook Choi, Sohn Young-Sun, "The Study on Dynamic Images Processing for Finger Languages", Korean Institute of Intelligent Systems, Vol 14, pp. 184-189, 2004.

[9] Frank Weichert, " Analysis of the Accuracy and Robustness of the Leap Motion Controller", Sensors, Vol. 13, pp. 6380-6393, 2013.

[10] Lawrence Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition." Proceedings of the IEEE, Vol 77, pp. 257-286, 1989.

[11] Yoon, H. S., Soh, J., Bae, Y. J., & Yang, H. S. "Hand gesture recognition using combined features of location, angle and velocity." Pattern recognition, Vol. 34, pp. 1491-1501, 2001.

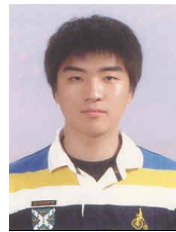
[12] Lee Hyung-Ji, Woo-Kyu Lee, Jae-Ho Chung, " Extracting feature vectors for hand gesture recognition of sign language", The Institute of Electronics and Information Engineers, Vol. 11, pp. 895-898, 1998.

[13] Nianjun Liu, Brian C. Lovel, Peter J,

Kootsookos, Richard I.A. Davis, "Model structure selection & training algorithms for an HMM gesture recognition system.", Frontiers in Handwriting Recognition, Ninth International Workshop on. IEEE, pp. 100-105, 2004.

[14] Oh Il-Seok, "Pattern Recognition", Kyobo library, pp. 349 - 354, 2008.

— 저 자 소 개 —



권 재 흥(정회원)  
2013년 경일대학교 컴퓨터과학과  
학사 졸업.  
2015년 중앙대학교 첨단영상대학원  
영상학과 석사 졸업.  
<주관심분야 : 컴퓨터 비전, 패턴  
인식>



김 태 용(평생회원)  
1986년 한양대학교 전기공학과  
학사 졸업.  
1988년 한양대학교 전자통신공학과  
석사 졸업.  
1998년 포항공과대학교 전자계산  
학과 박사 졸업.  
2003년~현재 중앙대학교 첨단영상대학원  
영상학과 교수  
<주관심분야 : 컴퓨터 비전, 영상처리, 컴퓨터 계  
임>