은닉 마르코프 모델 기반 손 제스처 적출을 위한 임계치 모델

(Threshold Model for Spotting Hand Gestures based on Hidden Markov Model)

이 현 규 † 김 호 연 †† 김 진 형 †††
(Hyeon-Kyu Lee) (Hoyon Kim) (Jin H. Kim)

요 약 본 논문에서는 실시간 제스처 인식 시스템을 위하여 연속적인 손동작으로부터 제스처 부분을 적출(spotting)하는 새로운 방법을 제안한다. 제안된 방법은 제스처의 구분문제(segmentation problem)를 해결할 수 있고 시공간적인 변이를 흡수할 수 있는 은닉 마르코프 모델에 기초를 두고 있다. 특히, 입력패턴으로부터 제스처가 아닌 패턴의 제거를 위하여, 임계치 모델(threshold model)이라는 새로운 모델을 도입하여 입력패턴의 임계 유사도를 계산하고 이를 이용하여 입력패턴이 제스처 패턴과 얼마나 유사한 지를 판정해 주도록 한다 제안된 방법을 이용하면 연속적인 손동작으로부터 93.38%의 신뢰도로 의미있는 제스처를 추출할 수 있다.

Abstract This paper proposes a new method of gesture spotting which extracts meaningful gestures from continuous hand motions for real-time gesture recognition systems. The proposed method is based on the HMM which can solve segmentation problem and can absorb spatio-temporal variability of gestures. To remove non-gesture patterns from input patterns, we introduce a new model called a threshold model that generates threshold likelihood of an input pattern and is used to qualify an input pattern as a gesture. The proposed approach has extracted meaningful gestures from continuous hand motions with 93.38% reliability.

1. 서 론

제스처(gesture)는 인간이 몸이나 손, 얼굴 등을 이용하여 의사를 표현하는 대화 방법이다. 이러한 여러 가지 제스처 중에서 손 제스처가 가장 표현력이 좋으며 자주 사용되고 있다[1, 2, 3]. 따라서 손 제스처는 여러 사람들에 의해 인간과 컴퓨터간의 새로운 인터페이스로서 연구되어 왔다[4, 5, 6]. 본 논문에서는 컴퓨터와 대화를 하려는 의도를 가진 손동작만을 제스처로 정의한다.

예측할 수 없는 입력신호로부터 의미 있는 부분을 추출하여 이를 인식하는 기법을 "패턴 적출(pattern spotting)" [7, 8]이라 한다. 제스처 인식은 제스처의 시작점과 끝점을 찾기 위한 방법을 필요로 한다는 면에서 패턴 적출 응용 중의 하나라고 볼 수 있다. 여기서부터는 제스처 패턴 추출을 이용한 제스처 인식을 "제스처 적출(gesture spotting)"이라 하기로 한다

제스처 적출을 위해서는 구분문제(segmentation problem)[9, 10]와 시공간적인 변이 문제[11, 12] 등의 두 가지 어려움을 해결해야 한다. 먼저, 제스처를 실시간에 인식하기 위해서는 제스처의 시작점과 끝점을 미리 결정하여 분리된 제스처 부분만을 인식기에 넘겨주어야 한다. 이처럼 제스처의 시작점과 끝점을 결정하는 것을 구분(segmentation)이라고 한다. 제스처를 취하는 사람이 한 제스처를 취하고 다음 제스처를 취하고자 할 때, 그의 손이 다음 제스처의 시작점까지 이동하면서 많은 점들을 지나고 인식기에서는 이러한 무의식적인 동작을 의도된 제스처로 인식하려 할 수 있기 때문에 구분문제는 제스처 인식기의 성능에 큰 영향을 주는 문제

† 비 회 원 한국과학기술원 전산학과
†† 학생회원 . 한국과학기술원 전산학과
††† 종신회원 : 한국과학기술원 전산학과 교수
논문접수 : 1997년 3월 25일
심사완료 1997년 9월 19일

라고 볼 수 있다. 구분문제를 해결하지 않으면, 인식기에서는 그림 1과 같이 입력신호의 모든 가능한 구간에 대해 매칭(matching)해 보아야 한다. 이때, 한 번의 매칭은 입력신호의 특정 구간에 대하여 모든 참조패턴과의 유사도를 계산하는 것이기 때문에 매칭의 수는 수행속도에 매우 큰 영향을 준다. 제스처 인식의 또 다른 어려움은 제스처를 행하는 사람마다 심지어는 같은 사람이 제스처를 반복할 때마다 그 모양과 동작시간이 다르다는 점이다. 이를 극복하기 위해서는, 인식기에서 시간적공간적 변이를 동시에 고려할 수 있어야 한다. 이상적으로는 인식기가 입력신호로부터 제스처 패턴을 추출해야 하며, 추출된 패턴이 시공간적인 변이를 포함하고 있더라도 참조패턴(reference pattern)과 매칭할 수 있어야 한다.

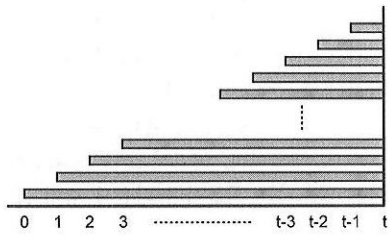


그림 1 현 시점(t)이 제스처의 끝점인 경우에 가능한 제스처 구간들

은닉 마르코프 모델(HMM; Hidden Markov Model)에서는 손동작의 대부분을 차지하는 비제스처 패턴(non-gesture pattern)에 대한 표현방법을 제공해 주고, 시공간적인 변이를 매우 잘 표현해 줄 수 있기 때문에, 본 논문에서는 제스처 적출을 위해 은닉 마르코프 모델을 적용하였다. 은닉 마르코프 모델은 시공간적인 변이를 가진 패턴을 표현하는데 가장 성공적으로 널리 쓰이는 방법[13]으로, 특히 온라인 필기인식[14, 15]과 음성인식[8, 16] 분야에서 성공적으로 이용되어 왔다.

그러나 은닉 마르코프 모델에서 제공하는 비제스처 패턴에 대한 표현방법을 그대로 사용하는 데에는 한계가 있다. 일반적으로 패턴 적출에서는 참조패턴에 대해서는 키워드 모델(keyword model)을 정의하고, 미지의 패턴에 대해서는 garbage 모델을 정의한다[16]. 이때, garbage 모델은 유한한 집합(문자 집합, 발음된 단어 집합)에 속하는 미지의 데이타를 이용하여 훈련한다. 그러나 제스처 적출에서는 비제스처 패턴을 정의할 수 없기 때문에, 훈련을 통하여 비제스처 패턴을 표현하는 garbage 모델을 이용할 수 없다. 이를 극복하기 위하여, 본 논문에서는 은닉 마르코프 모델의 내재적 구분(internal segmentation) 속성을 응용하여, 훈련된 제스처 모델의 상태들로 구성되었으며 제스처 모델의 매칭 결과의 수준을 판정하기 위해 사용되는 새로운 모델을 도입한다. 내재적 구분 속성은, 훈련된 은닉 마르코프 모델에서 각 상태(state)와 전이(transition)는 내재적으로 각각 제스처 패턴의 부패턴(sub-pattern)과 그들의 순서를 표현한다는 속성이다. 이 속성을 이용하면, 제스처의 부패턴의 조합순서를 바꾸어 만든 새로운 패턴과 매칭될 수 있는 또 다른 모델을 구축할 수 있다. 또한, 제스처 모델의 상태를 이용하여 모든 상태가 상호 연결된 ergodic 모델을 구축하면, 제스처의 부패턴이 어떠한 순서로 조합되어 있더라도 매칭될 수 있는 모델을 구축할 수 있다.

훈련된 제스처 모델의 상태를 이용함으로써 새로운 모델이 참조패턴의 모든 부패턴을 표시할 수 있으며, 이러한 상태들을 이용하여 ergodic 모델을 구축함으로써 새로운 모델이 참조패턴의 부패턴이 어떠한 순서로 조합되어 있더라도 잘 매칭될 수 있도록 한다. 그러나, 새로운 모델의 각 상태에서는 타 상태로의 전이(outgoing transition)의 수가 원래 해당하는 제스처 모델에서의 전이의 수보다 많기 때문에, 제스처 패턴은 제스처 모델에 좀 더 잘 매칭된다. 이러한 새로운 모델의 출력은 제스처 모델의 출력에 대한 적응형 임계치(adaptive threshold)로 사용될 수 있으므로, 새로운 모델을 임계치 모델(threshold model)이라 부르기로 한다. Garbage 모델의 출력은 입력패턴과 garbage 패턴과의 유사도 점검에 사용되지만, 임계치 모델의 출력은 입력패턴과 제스처 패턴의 유사도에 대한 평가자료로 활용된다는 점에서 garbage 모델과 구별된다.

은닉 마르코프 모델에 기반을 둔 임계치 모델의 성능평가 실험을 위하여, 제스처를 사용하여 PowerPoint™의 화면이동을 제어할 수 있는 PowerGesture 시스템을 구축하였다[17]. 제안된 방법을 이용하여 손동작 중에서 제스처를 적출하는 실험은 93.38%의 신뢰도를 보여주었으며, 실험 시스템의 평균 적출속도는 4.45 프레임/초를 나타냈다.

본 논문에서는 임계치 모델을 이용한 은닉 마르코프 모델 기반 제스처 적출방법을 제안하고 그 성능을 평가하고자 한다. 본 논문의 구성은 다음과 같다. 제 2장에서는 패턴 적출의 세 가지 주요 방법들을 비교 검토하고, 본 논문에서 채택한 은닉 마르코프 모델에 대하여 설명한다. 제 3장에서는 임계치 모델과 이를 이용한 제스처 적출 네트워크에 대해 소개하고, 제스처의 시작점

과 끝점을 결정하는 끝점탐색기를 설명한다. 제 4장에서는 제스처 변별력 평가와 제스처 적출 평가로 구분하여 실험 결과를 설명하고, 마지막으로 제 5장에서는 결론을 짓는다.

2. 관련 연구

2.1 패턴 적출 방법론 비교

시간 및 공간적인 변이를 가진 패턴을 추출하는 기법으로는 DTW(Dynamic Time Warping)[12, 19], 신경망[3, 6], 은닉 마르코프 모델(Hidden Markov Model)[11, 13, 18] 등을 들 수 있다.

DTW는 입력패턴과 참조패턴 사이의 거리를 계산하기 때문에 여러 개의 후보를 낼 수 있다. 또한, DTW는 대표적인 모양의 템플리트를 참조패턴으로 사용하기 때문에 훈련 데이타가 적은 환경에서도 사용이 가능하다[12]. 그러나 DTW는 매칭과정에서 변이정보를 계산해야 하기 때문에 공간적인 변이를 처리하기 위해서는 추가적인 참조패턴이 필요하며, 이는 매칭시간의 증가를 유발한다. 또한 DTW는 미지의 패턴을 표현하기 위한 방법론을 갖고 있지 못하다.

신경망은 입력패턴의 사후확률(posterior probability)을 계산하기 때문에 한 개의 후보밖에 낼 수 없으며, 미지의 패턴에 대한 표현 방법이 없다. 신경망의 훈련을 위해서는 방대한 양의 훈련 데이타를 필요로 하기 때문에 데이타를 구하기가 용이하지 않은 응용에는 부적합하다. 재귀 신경망(recurrent neural network)이나 시간지연 신경망(time delayed neural network)을 이용하여 참조패턴의 시간적인 변이를 고려할 수는 있으나, 일반적으로 신경망을 이용하여 시간적인 정보를 표현하기는 어렵다. 따라서, 신경망은 대개 포스처(posture)라는 정적인 제스처를 인식하는데 사용되고 있다[3].

은닉 마르코프 모델은 입력패턴과 참조패턴 사이의 유사도를 평가하기 때문에 여러 개의 후보를 낼 수 있으며, 시공간적인 변이가 각 상태와 전이에 확률적으로 표현되어 있기 때문에 매칭과정에서 참조패턴의 시공간적인 변이에 대하여 추가적으로 고려할 필요가 없다. 또한, 미지의 패턴이 유한하다면 이를 훈련시켜 미지의 패턴을 표현하는 garbage 모델을 구축할 수 있다[16].

2.2 은닉 마르코프 모델

은닉 마르코프 모델은 시공간적인 변이를 가진 사건을 모델링 하는데 성공적으로 널리 사용되어 왔으며, 특히 음성인식과 온라인 필기인식 분야에서 성공적으로 응용되어 왔다. 이 기법은 시공간적인 정보를 자연스럽게 모델링 할 수 있으며 학습과 인식을 위한 효과적이고 우수한 알고리즘을 가지고 있기 때문에 여러 분야에서 성공적으로 응용되고 있다[13].

은닉 마르코프 모델이란 전이(transition)에 의해 연결된 상태(state)들의 모임으로 각 전이는 두 가지 확률의 집합을 표현한다. 하나는 전이를 하기 위해 필요한 전이확률(transition probability)이고, 다른 하나는 전이가 발생할 때마다 유한한 알파벳에 속하는 각 출력심볼(output symbol)을 발생시키는 조건부 확률을 나타내는 출력확률(output probability)이다[13, 15]. 좀 더 수학적으로 정리해 보면, 은닉 마르코프 모델은 다음의 요소에 의해 정의된다.

- $\{s_1, s_2, s_3, \cdots, s_N\}$ — N 상태의 집합이다. 이때, 시간 t에서의 상태를 확률변수 q_t로 표시한다.
- $\{v_1, v_2, v_3, \cdots, v_M\}$ — 서로 다른 M 관측심볼의 집합으로서 사용 가능한 알파벳을 나타낸다. 이때, 시간 t에서의 관측심볼을 확률변수 O_t로 나타낸다. 관측심볼은 모델링을 하고자 하는 시스템의 실질적인 출력에 해당한다.
- $A = \{ a_{ij} \}$ — 상태전이 확률분포를 표현하기 위한 $N \times N$ 행렬이다. 여기서 a_{ij}는 상태 s_i에서 s_j로 전이를 일으키는 확률이며 다음과 같이 표시한다.

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

- $B = \{ b_j(k) \}$ — 관측심볼의 확률분포를 표현하기 위한 $N \times M$ 행렬이다. 여기서 $b_j(k)$는 상태 s_j에서 시간 t에 v_k 관측심볼을 보이게 될 확률이며 다음과 같이 표시한다.

$$b_j(k) = P(O_t=v_k \mid q_t=s_j)$$

- $\pi = \{ \pi_i \}$ — 초기 확률분포의 집합이다. 여기서, π_i는 s_i가 초기상태가 될 확률이며 다음과 같이 표시한다.

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

여기서 A, B 및 π는 확률변수이므로 다음의 속성을 만족하여야 한다.

- $\Sigma_i a_{ij} = 1 \quad \forall i, \ a_{ij} \geq 0$
- $\Sigma_k b_j(k) = 1 \quad \forall j, \ b_j(k) \geq 0$
- $\Sigma_i \pi_i = 1 \quad \pi_i \geq 0.$

은닉 마르코프 모델은 대개 위의 다섯 요소들 중에서 확률변수만을 포함하여 $\lambda = \{ A, B, \pi \}$의 형태로 표시된다.

은닉 마르코프 모델에서는 매칭이 이뤄진 관측심볼들에 대해 모델의 어떤 상태들을 거쳐 결과가 나왔는지를

추적하는 방법을 제공하는데, 이를 Viterbi 알고리즘이라고 하고 추적된 상태열을 Viterbi 경로라고 한다[13]. Viterbi 알고리즘은 "초기 상태와 초기의 결정이 어떻든 간에 남아 있는 결정들은 첫번째 결정으로부터 나온 상태에 대해 최적이어야 한다"[20]는 최적의 원리(principle of optimality)에 근거를 두고 있다.

2.3 제스처 적출을 위한 은닉 마르코프 모델

은닉 마르코프 모델은 시간적인 구조를 가지고 있어서 음성과 온라인 필기를 위한 자연스러운 표현방법이 될 수 있으며, 특성벡터를 이용하여 제스처를 표현할 수도 있다. 즉, 제스처를 특성벡터들의 시간순서에 따른 관측열로 표현할 수 있다. 여기서 특성벡터는 손의 이동방향(sin 및 cos값)으로 구성된다. 본 논문에서는 이산 은닉 마르코프 모델을 사용하므로, 특성벡터는 벡터 양자화기에 의하여 16개의 코드워드 중의 하나로 변경된다.

제스처 적출기는 손동작을 계속적으로 관찰하다가 제스처가 나타나면 이를 적출해야 하므로 손동작에 대한 모델링이 가능해야 한다. 이때 적출의 복잡성을 줄이기 위하여 제스처는 입력패턴의 마지막에만 존재한다는 가정을 하면 다음과 같이 손동작을 표현할 수 있다.

$$\text{손동작} ::= \{\text{비제스처}\}^* \bullet \{\text{제스처}\}$$

여기서 '*'은 없거나 1번 이상 반복됨을 의미한다. 비제스처는 제스처와 같이 특성벡터에 의해 표현되지만 제스처와는 달리 시간적인 순서의 표현이 불필요하다. 이것은 제스처와는 관계없는 임의의 손동작이 반복되다가 어느 순간 제스처가 발생하면, 즉시 해당 제스처를 적출할 수 있는 가능을 필요로 한다는 것을 말하고 있다.

은닉 마르코프 모델의 구조를 선택하는 것은 훈련 데이타의 양이나 표현하려고 하는 모델에 따라 다르다. 제스처는 시간적 순서에 대한 고려가 필수적이므로 좌-우 모델(left-right model)을 사용하는 것이 좋다. 좌-우 모델에서는 시간의 흐름에 따라 상태(state)의 번호가 같거나 증가하는 방향으로 진행되는 특성을 가지고 있다. 즉, 상태가 왼쪽에서 오른쪽으로 전이한다 (그림 2). 제스처 모델의 상태의 수는 제스처의 복잡성에 따라 5 ~ 8개로 설정한다.

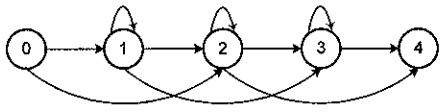
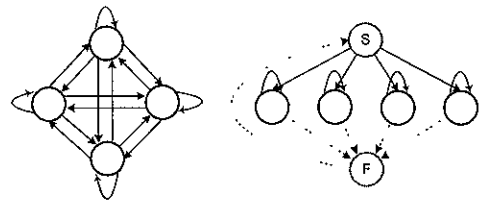


그림 2 제스처 모델

그러나 임계치 모델은 제스처를 포함한 모든 손 동작을 표현할 수 있어야 하기 때문에 ergodic 모델을 사용한다 (그림 3(a)). Ergodic 모델은 각 상태가 다른 모든 상태로부터 한번의 전이에 의해 도달할 수 있는 완전히 연결된 모델로서, 참조모델의 부패턴을 이용한 모든 조합결과에 대해서도 잘 매칭될 수 있도록 구성하는데 유용하다. 상태의 수가 많은 경우, ergodic 모델을 표현하려면 모든 상태들을 상호 연결해야 한다. 본 논문에서는 null 상태와 null 전이를 도입하여 ergodic 모델과 같은 기능의 단순화된 구조를 사용하였다 (그림 3(b)). Null 상태는 관측심볼을 처리하지 않는 상태이며, null 전이는 전이에 따르는 확률계산이 불필요한 전이를 의미한다.



(a) 원래의 모델 구조 (b) 단순화된 모델 구조

그림 3 Ergodic 모델의 구조. S, F는 null 상태이고, 점선은 null 전이를 의미한다.

3. 임계치 모델을 이용한 제스처 적출

3.1 임계치 모델

은닉 마르코프 모델의 내재적 구분(internal segmentation) 속성이란 훈련된 모델의 각 상태와 자체전이는 제스처 패턴의 부패턴을 표현하고, 타 상태로의 전이는 제스처 패턴 내에서 부패턴의 조합순서를 표현하고 있다는 속성이다. 이 속성을 이용하면, 부패턴의 순서에 관계없이 원래의 모델과 비슷한 유사도를 출력하는 새로운 모델을 구축할 수 있다.

본 논문에서는 제스처 모델을 5 ~ 8개의 상태를 가진 좌-우 모델로 구축하였고, Baum-Welch 알고리즘을 이용하여 각 모델을 훈련하였다. 훈련된 모든 제스처 모델에 대하여 각 모델 내의 상태(state)와 자체전이(self-transition)를 추출한 후, 모든 상태들이 상호 연결된 ergodic 모델을 새로이 구축하였다. 새로운 모델에서 각 상태는 한 번의 전이에 의해 타 상태에 도달할 수 있다. 또한, 새로운 모델에서 각 상태와 자체전이의 확률은 제스처 모델에서의 확률을 그대로 보유하고 있으며, 타 상태로의 전이확률(outgoing transition pro-

bability)은 각 상태에서의 전이확률의 합이 1이라는 사실로부터 자체전이 확률을 이용하여 계산하며, 모든 전이에 대해 동일한 값으로 할당한다. 그림 4는 ST와 FT라는 두 개의 null 상태를 가지고 있는 임계치 모델을 보여준다. Null 상태는 입력(observation)에 대한 처리를 하지 않는 상태이다

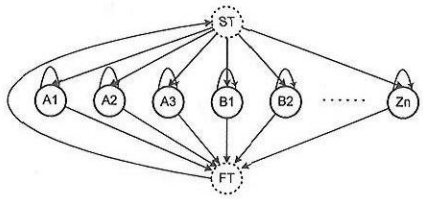


그림 4 임계치 모델의 구조

각 상태와 자체전이의 확률을 유지함으로써 새로운 모델이 참조패턴의 모든 부패턴을 표현할 수 있도록 하였으며, ergodic 모델을 구축함으로써 참조패턴의 부패턴을 이용한 모든 조합결과에 대해서도 잘 매칭될 수 있도록 하였다. 그러나, 새로운 모델에서는 타 상태로의 전이확률이 제스처 모델에서보다 작기 때문에 제스처 패턴은 제스처 모델에 더 잘 매칭된다. 따라서, 새로운 모델의 출력은 제스처 모델의 출력에 대한 적응형 임계치(adaptive threshold)로 사용될 수 있다. 이러한 이유로 새로운 모델을 임계치 모델(threshold model)이라 하였다.

3.2 제스처 적출 네트워크

제스처 모델을 훈련시키고 임계치 모델이 구축되면, 연속적인 손동작으로부터 제스처를 적출 하기 위하여 각각의 제스처 모델들과 임계치 모델을 연결한 제스처 적출 네트워크(GSN; Gesture Spotting Network)을 구축한다(그림 5). 그림에서 S는 null 시작상태를 의미한다. 그림에서 이름표가 붙어 있는 각각이 하나의 제스처 모델을 나타내며, 마지막에 ergodic 모델로 표현된 것이 임계치 모델이다.

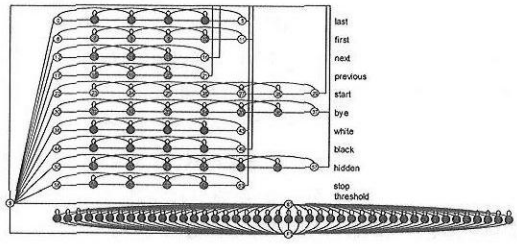


그림 5 제스처 적출 네트워크

신뢰성 있는 적출을 위하여, 모델 전이확률은 다음과 같이 되도록 조절한다.

$$P(X_G \mid \lambda_G)\ p(G) > P(X_G \mid \lambda_{TM})\ p(TM) \quad (1)$$

$$P(X_{TM} \mid \lambda_G)\ p(G) < P(X_{TM} \mid \lambda_{TM})\ p(TM) \quad (2)$$

이때, X_G, X_{TM}, λ_G, λ_{TM}은 각각 제스처 패턴, 비제스처 패턴, 올바른 제스처 모델, 임계치 모델을 표시한다. 또한, $p(G)$와 $p(TM)$는 각각 제스처 모델과 임계치 모델로의 모델 전이확률을 표시한다. (식 1)은 제스처 패턴은 제스처 모델에 가장 잘 매칭되어야 함을 표시하며, (식 2)는 비제스처 패턴은 제스처 모델에 매칭되지 않아야 한다는 것을 표현하고 있다. 본 논문에서는, $p(TM)$의 변화에 따른 적출결과에 따라 $p(TM)$값을 조율한다.

제스처 모델의 유사도가 임계치 모델보다 높아지면, 그 시간이 바로 제스처의 후보끝점이 된다. 제스처의 시작점은 Viterbi 경로를 역추적하면 쉽게 구할 수 있다. 이것은 제스처 모델이 좌-우 모델이므로 시작상태를 거치지 않고는 끝상태에 갈 수 없기 때문이다.

3.3 끝점탐색기

개별 모델의 시간 경과에 따른 유사도의 변화를 보면, 대개는 임계치 모델의 유사도가 가장 좋게 나온다. 그러나 현시점이 제스처의 끝점 근처에 이르게 되면, 그림 6의 (a)에서 보는 바와 같이 해당하는 제스처 모델의 유사도가 좋아진다(시간=12). 그림에서, 시간이 12에 이르기 전까지는 *last* 모델의 유사도가 임계치(그림에서

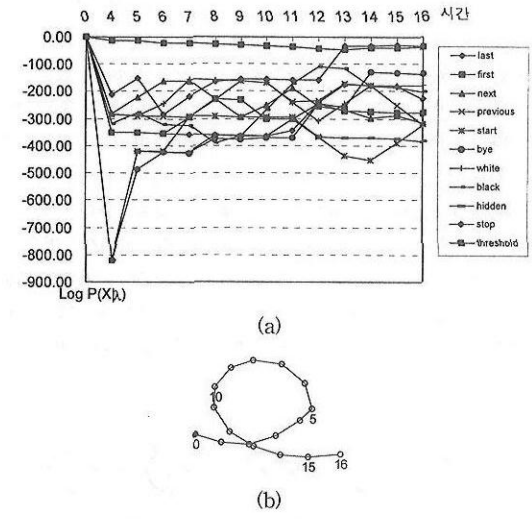


그림 6 (a) 제스처 적출 네트워크의 log 유사도 그래프 (b) 입력패턴의 궤적

*threshold*로 표기) 모델의 유사도보다 작게 나타나지만, 시간이 12인 시점부터는 *last* 모델의 유사도가 임계치 모델의 유사도보다 높게 나타난다. 시간이 12인 점부터 16인 점 사이의 모든 시간이 *last* 제스처의 후보끝점이 된다. 각 후보끝점들의 시작점은 Viterbi 경로를 역추적하여 찾는다.

끝점탐색기는 후보끝점들 중에서 최적의 끝점을 찾는 알고리즘으로 다음과 같은 경우에 동작을 시작한다. 먼저, 하나의 제스처를 찾은 후 그 다음 제스처의 마지막 후보끝점을 찾았을 경우이고, 다음은 찾은 제스처의 마지막 후보끝점 이후에 일정한 시간이 지나도록 다른 제스처가 나타나지 않았을 경우이다.

끝점탐색기의 탐색기준(detection criteria)은 현재 고려 중인 제스처 바로 다음에 이어지는 패턴에 대한 heuristic에 의해 정의되며 다음과 같다. 여기서, 표현상의 편의를 위하여 현재 끝점을 결정하고자 하는 제스처를 '현 제스처'로, 그 다음에 이어지는 제스처를 '다음 제스처'로, 현 제스처의 첫번째 후보끝점을 '첫 후보'로, 현 제스처의 마지막 후보끝점을 '마지막 후보'로 표시한다.

(1) 그림 7의 (a)와 같이 현 제스처 다음에 이어지는 패턴이 제스처가 아닌 경우, 마지막 후보를 끝점으로 선택한다.

(2) 현 제스처 다음에 이어지는 패턴이 제스처인 경우, 다음의 두 가지 선택이 있다.

(a) 그림 7의 (b)와 같이 다음 제스처의 시작점이 첫 후보보다 앞에 있으면, 다음 제스처가 현 제스처의 끝점을 포함하여 펼쳐지는 것이기 때문에 현 제스처를 다음 제스처의 일부라고 간주하여 현 제스처의 모든 후보끝점들을 무시한다.

(b) 그림 7의 (c)와 같이 다음 제스처의 시작점이 첫 후보와 마지막 후보 사이에 놓여 있으면, 마지막 후보를 끝점으로 선택한다.

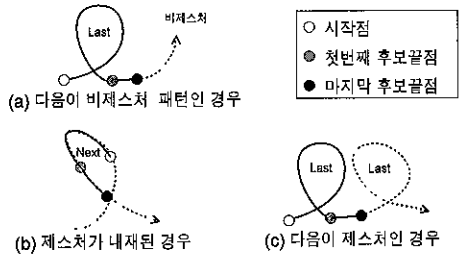


그림 7 끝점 결정 방법

또한, 끝점탐색기는 사용자의 즉각적인 처리의도를 파악하기 위한 고려를 해야 한다. 제스처에 대한 즉시처리를 요구하는 사용자의 의도로는 손이 카메라 영역을 벗어나거나 손이 한 곳에서 일정시간 동안 움직이지 않는 경우가 있다. 일단 사용자의 즉시처리 요구가 발견되면, 적출을 멈추고 현재까지 발견된 모든 제스처의 끝점을 결정한다.

4. 실험 및 결과

은닉 마르코프 모델에 기반을 둔 임계치 모델의 성능 평가를 위하여 제스처를 명령어를 이용하여 PowerPoint™ 화면이동을 제어할 수 있는 PowerGesture 시스템을 구축하였다. 또한, PowerPoint™를 이용한 연설 과정에서 사용되는 명령어에 대해 유사한 의미를 가진 제스처를 할당하였다(표 1).

표 1 실험에 사용한 제스처

모 양	이 름	의 미
	last	마지막 화면으로 이동하기
	first	첫 화면으로 이동하기
	next	다음 화면으로 이동하기
	previous	이전 화면으로 이동하기
	start	연설모드 시작하기
	bye	프로그램 종료
	white	화면을 백색으로 바꾸기
	black	화면을 흑색으로 바꾸기
	hidden	숨은 화면 보여주기
	stop	연설모드 끝내기

실험에 사용한 시스템은 연설지원용으로 PowerGesture 시스템으로 명명하였으며, 카메라로부터 손동작에 대한 영상 프레임을 얻고 제안된 적출방법을 이용하여 손동작으로부터 제스처를 인식한 후, 인식된 명령어를 연설자료 화면제어에 사용하는 제스처 인터페이스를 갖고 있다(그림 8). PowerGesture 시스템은 Windows 95를 운영체제로 하는 Pentium Pro PC에서 구축하였다. 카메라에서는 영상 프레임을 얻으면, 손 추적기에서는 프레임 내에 있는 손을 찾는다. 백터 양자화기에서는 이렇게 찾은 손의 위치정보와 이전에 얻은 손의 위치정보를 이용하여 방향 특성(directional feature)을 추출하고, 방향 특성은 은닉 마르코프 모델의 입력이 되는 16

개의 codeword 중의 하나로 양자화된다. 제스처 적출기에서는 codeword를 이용하여 은닉 마르코프 모델에 대한 매칭을 수행하여 정의된 제스처가 입력되었는지를 확인하고, 있으면 이를 적출한다. 적출된 제스처는 PowerPoint™의 명령어로 변경되어 연설자료의 화면을 제어한다.

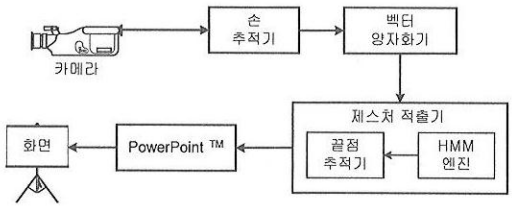
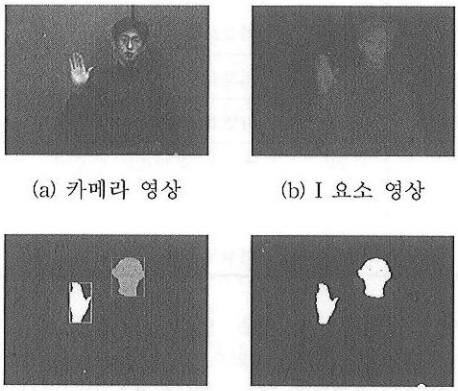


그림 8 Power Gesture 시스템의 구성도

손동작 추적기는 YIQ 색상체계의 I 요소가 피부색에 민감하게 반응하기 때문에 YIQ 색상체계를 이용하였으며, 임계치를 이용하여 I 요소를 이진영상으로 변환한다. 이렇게 만들어진 이진영상에 대하여 one-pass 레이블링 알고리즘[21]을 이용하여 물체구분(object segmentation)을 하고 손을 찾는다. 영상처리 과정의 단순화를 위하여 단순배경에서 오른손만을 이용하였다 (그림 9).



(a) 카메라 영상 (b) I 요소 영상

(d) 손추적 결과 (c) 이진 영상

그림 9 손 추적 과정

제스처 적출을 평가하는 방법으로는 두 가지 방법이 있다. 첫째 방법은 제스처 모델의 훈련 정도와 임계치 모델이 적절한 임계치를 출력하는 지를 평가하기 위한 제스처 변별력 평가이고, 둘째 방법은 제스처 적출기가 제스처를 얼마나 잘 구분해 내는가를 평가하는 제스처 적출 평가이다. 본 논문에서는 위의 두 가지 방법을 모두 이용하여 제스처 적출능력을 평가하였다.

4.1 제스처 변별력 평가

표 2과 같이 여섯 명의 피실험자로부터 2,500개의 독립된 제스처를 얻어 제스처 모델의 훈련 및 평가 데이타로 사용하였다. 제스처 적출의 성공여부는 제스처 모델과 임계치 모델의 변별력에 크게 의존하기 때문에, 모델간의 변별력 평가를 위하여 독립된 제스처에 대한 인식을 시도하였다. 표 3과 같이, 이 평가에서는 제스처 간의 유사성이 높음에도 불구하고 변별력이 강하게 나타나고 있음을 알 수 있다(인식률=98.19%). 이 평가에서 나타난 대부분의 에러는 제스처 모델의 유사도가 임계치 모델보다 높지 않으면 제스처로서의 자격을 부여하지 않는 이유로 나타난 것으로 제스처 모양의 왜곡이 심한 경우에 발생하였다.

표 2 제스처 모델용 훈련 데이타

	훈련 데이타	평가 데이타
last	196	54
first	195	55
next	198	52
previous	195	55
start	202	48
bye	203	47
white	196	54
black	202	48
hidden	212	38
stop	205	45
합 계	2,004	496

표 3 제스처 변별력 평가결과

	평가 데이타	정인식	오인식	인식률(%)
last	54	54	0	100.00
first	55	54	1	98.18
next	52	51	1	98.08
previous	55	55	0	100.00
start	48	46	2	95.83
bye	47	45	2	95.74
white	54	53	1	96.30
black	48	46	2	95.83
hidden	38	38	0	100.00
stop	45	45	0	100.00
합 계	496	487	9	98.19

4.2 제스처 적출 평가

두 번째 평가는 앞의 평가에서 사용한 제스처 모델과

임계치 모델을 이용하여 제안된 방법의 적출능력을 평가하기 위한 것이다. 이를 위해 피실험자 1인으로부터 60개의 평가 데이타를 얻었다. 각 평가 데이타는 1개 이상의 제스처를 포함하고 있으며 200 프레임으로 구성되었다.

제스처 적출에는 제스처가 아닌 패턴을 제스처라고 하는 삽입에러(insertion error), 제스처를 찾지 못하는 삭제에러(deletion error), 다른 제스처와 혼동하는 대치에러(substitution error)와 같은 세 가지 형태의 에러가 나타난다. 삽입에러는 (식 3)에 나타난 바와 같이 발견률(detection ratio)을 산출하는데 고려되지 않는다. 그러나, 삽입에러로 인해 끝점탐색기가 입력으로부터 실제 제스처의 일부 또는 전체를 제거하는 우를 범할 수 있기 때문에 삽입에러는 삭제 또는 대치에러를 야기시키는 원인이 될 수 있다. 따라서, 제스처 적출의 성능을 평가하기 위해서는 발견률과는 다른 척도가 필요하다. 본 논문에서는 삽입에러를 고려한 신뢰도(reliability)를 성능평가의 척도로 삼았다 (식 4).

$$\text{발견률} = \frac{\text{정인식된제스처}}{\text{총제스처}} \tag{3}$$

$$\text{신뢰도} = \frac{\text{정인식된제스처}}{\text{총제스처} + \text{삽입에러}} \tag{4}$$

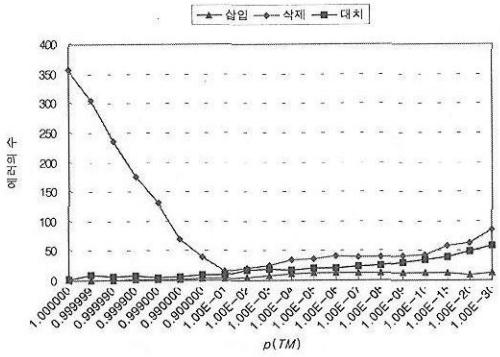


그림 10 임계치 모델로의 모델전이 확률에 따른 에러의 수

또한, 실험에서는 임계치 모델로의 모델 전이확률($p(TM)$)을 바꾸어 가면서 에러를 측정하였다(그림 10). $p(TM)$값이 크면 임계치 모델의 유사도가 높아져 많은 제스처가 제스처로서의 자격을 상실하게 되고, 반대로 낮으면 임계치 모델의 유사도가 낮아져 제스처 뿐만 아니라 제스처와 약간 비슷하게 생긴 패턴까지도 검출된다. 그림에서 보듯이 $p(TM)$이 1.0에서 0.1로 감소할수록 삭제에러가 급격히 감소하지만, $p(TM)$이 0.1을 지나면서부터는 삭제에러는 서서히 증가한다. 기대했던 결과는 $p(TM)$이 특정값(여기서는 0.1) 이하로 작아지면 삭제에러에 변화가 없는 것이었으나, 실제로는 삭제에러가 조금씩 증가한다. 이것은 삽입에러의 증가가 삭제에러에 영향을 주기 때문이다. 삭제에러는 발견률에 직접적인 영향을 주지만 삽입에러는 그렇지 않다. 그러나 일부 삽입에러가 삭제 또는 대치에러를 야기시키기 때문에 삽입에러가 삭제에러에 완전히 독립이라고는 말할 수 없다. 표 4은 가장 좋은 신뢰도를 보인 $p(TM)$이 0.1인 경우의 실험결과를 보여주고 있다.

표 4 $p(TM)$ = 0.1인 경우의 적출결과

	제스처의 수	결 과				
		삽입에러	삭제에러	대치에러	정인식	신뢰도
last	41	1	1	2	38	90.48
first	48	1	2	2	44	89.80
next	57	1	0	1	56	96.55
previous	41	0	4	1	36	87.80
start	28	0	3	0	25	89.29
bye	35	0	2	1	32	91.43
white	45	0	1	0	44	97.78
black	49	0	2	1	46	93.88
hidden	39	0	1	1	37	94.87
stop	37	0	0	0	37	100.00
합 계	420	3	16	9	395	93.38

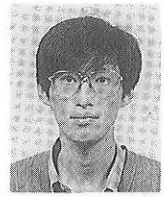
5. 결 론

본 논문에서는 카메라를 이용하여 얻어진 화상계열에 대하여 임계치 모델을 이용한 제스처 적출이 좋은 결과를 내고 있음을 보였다. 제안된 방법은 평균 초당 4.45 프레임을 처리할 수 있으며, 손동작에 제약을 주지 않고도 93.38%의 신뢰도로 제스처를 적출할 수 있음을 보였다. 실험결과는 임계치 모델이 단순하면서도 입력패턴이 제스처와 얼마나 유사한 지를 판정하는 데 매우 효과적임을 보였다. 그러나 제스처 모델의 수의 증가에 따라 임계치 모델의 상태의 수가 증가하여 적출속도가 느려진다. 따라서 제스처 모델의 수의 증가와 관계없이 임계치 모델의 상태의 수를 일정한 수준으로 유지시키는 것이 향후 연구해야 할 과제이다.

참 고 문 헌

[1] S.Fels and G.E.Hinton, "Glove-talk: A Neural

Network Interface between a Dataglove and a Speech Synthesizer", IEEE transactions on Neural Networks, Vol. 4, pp. 2-8, 1993.
[2] C.Maggioni, "A Novel Gestural Input Device for Virtual Reality", Proc. of IEEE Virtual Reality Annual International Symposium, pp. 118-124, 1993.
[3] K.Vaananen and K.Boehm, "Gesture Driven Interaction as a Human Factor in Virtual Environments - An Approach with Neural Networks", Chapter 7 of Virtual Reality Systems, R. Earnshaw, M. Gigante, H. Jones (eds.), Academic Press, pp. 93-106, 1993.
[4] F.Quek, "Toward a Vision-based Hand Gesture Interface", Proc. Virtual Reality Software and Technology Conference (VRST), pp. 17-31, 1994.
[5] W.T. Freeman and C.D.Weissman, "Television control by hand gestures", Proc. Int. Workshop on Automatic Face- and Gesture-Recognition (IWAFGR), pp. 179-183, 1995.
[6] R.Kjeldsen and J.Kender, "Visual Hand Gesture Recognition for Window System Control", Proc. Int. Workshop on Automatic Face- and Gesture-Recognition (IWAFGR), pp. 184-188, 1995.
[7] F.R.Chen, L.D.Wilcox and D.S.Bloomberg, "Word Spotting in Scanned Images using Hidden Markov Models", Proc. Int. Conference on Acoustics, Speech and Signal processing (ICASSP), Vol. V, pp. 1-4, 1993.
[8] R.C.Rose, "Discriminant Wordspotting Techniques for Rejection Non-vocabulary utterances in Unconstrained Speech", Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Vol II, pp. 105-108, 1992.
[9] T.Baudel and M.Beaudouin-Lafon, "CHARADE: Remote Control of Objects using Free-Hand Gestures", Communications of ACM, Vol. 36, No. 7, pp. 28-35, 1993.
[10] A.Wexelblat, "Natural Gesture in Virtual Environments", Proc. Virtual Reality Software and Technology Conference (VRST), pp. 5-16, 1994.
[11] T.Starner and A.Pentland, Real-Time American Sign Language Recognition from Video Using Hidden Markov Models, Technical Report TR-375, Media Lab, MIT, 1995.
[12] K.Takahashi, S.Seki, and R.Oka, "Spotting Recognition of Human Gestures from Motion Images", Technical Report IE92-134, The Institute of Electronics, Information and Communication Engineers(Japan), pp. 9-16, 1992.
[13] X.D.Huang, Y.Ariki, and M.A.Jack, Hidden Markov Models for Speech Recognition, Edinburgh Univ. Press, 1990.
[14] S.Lee, H.Lee, and J.Kim, "On-Line Cursive Script Recognition Using an Island-Driven Search Technique", Proc. Int. Conference on Document Analysis and Recognition (ICDAR), pp. 886-889, 1995.
[15] B.Shin, An HMM-Based Statistical Framework For Modeling On-line Cursive Script, Ph.D. Thesis, 1995.
[16] L.D.Wilcox, and M.A.Bush, "Training and Search Algorithms for an Interactive Wordspotting System", Proc. Int. Conference on Acoustics, Speech and Signal processing (ICASSP), Vol II, pp. 97-100, 1992.
[17] 이현규, 김진형, "PowerGesture: 제스처 Spotting 기법을 이용한 발표 지원 시스템", HCI '97 학술대회 발표논문집, pp. 90-95, 1997.
[18] J.Yamato, H.Ohya and K.Ishii, "Recognizing Human Action in Time-Sequential Images using Hidden Markov Model", Proc. 1992 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 379-385, 1992.
[19] S.Seki, K.Takahashi and R.Oka, "Gesture Recognition from Motion Images by Spotting Algorithm", Proc. of Asia Conference on Computer Vision (ACCV), 1993.
[20] D.A.Pierre, Optimization Theory with Applications, Dover Edition, 1986.
[21] 고일주, 최형일, "영역 추출과 추적에 의한 손 영역 획득", 한국정보과학회 '96 봄학술논문집, pp. 239-242, 1996.



이 현 규

1985년 서울대학교 전자계산기공학과(현 컴퓨터공학과) 학사. 1987년 한국과학기술원 전산학과 석사. 1987년 ~ 1990년 한국전기통신공사 품질보증단 전임연구원. 1990년 ~ 1991년 휴먼컴퓨터(주) 선임연구원. 1991년 ~ 현재 (주) 핸디소프트 기술이사. 1994년 10월 신소프트웨어 상품대상 10월의 대상 개발자 상 수상. 1994년 12월 신소프트웨어 상품대상 1994년 대상 개발자 상 수상. 1995년 ~ 현재 한국과학기술원 전산학과 박사과정. 관심분야는 전문가시스템, 패턴인식, 영상처리, 컴퓨터 비전, HCI 등.

김 호 연

1992년 연세대학교 전산학과 학사. 1994년 한국과학기술원 전산학과 석사. 1997년 1월 ~ 2월 일본 NHK 放送技術研究所방문연구원. 1994년 ~ 현재 한국과학기술원 전산학과 박사과정. 관심분야는 패턴인식, 문자인식, NMM, 영상처리, 신경망, 기계학습 등.

김 진 형

1971년 서울대학교 공과대학 졸업. 1979년 UCLA 전산학 석사학위 취득. 1983년 UCLA 전산학 박사학위 취득. 1973년 ~ 1976년 KIST 전산실 연구원. 1976년 ~ 1977년 미국 캘리포니아 도로국 연구원. 1981년 ~ 1985년 미국 Hughes 인공지능 센터 선임 연구원. 1986년 ~ 1988년 한국 정보과학회 산사 인공지능 연구회 위원장. 1989년 ~ 1990년 미국 IBM 와트슨 연구소 초빙 과학자. 1985년 ~ 현재 한국과학기술원 전산학과 교수. 1991년 ~ 현재 한국과학기술원 전산학과 교수, 1991년 ~ 현재 인공지능 연구센터 부소장. 1995년 ~ 현재 한국연구개발정보센터 소장. 관심분야는 인식 시스템, 지능형 인터페이스, Computer-Aided Education, Groupware, 지식기반 시스템임.