

테이블 탑 디스플레이에서 사용자 독립적인 온라인 필기 숫자 인식

김지웅, 김의철, 김수형
전남대학교 전자컴퓨터공학부
e-mail: haesanjin2@naver.com

User Independent On-line Handwritten Numeral Recognition in Table Top Display

Ji-Woong Kim, Eui-Chul Kim, Soo-Hyung Kim
Department of Electronics and Computer Engineering, Chonnam National University

요 약

테이블 탑 디스플레이는 사람에게 친숙한 상호작용의 수단인 손을 인터페이스 수단으로 이용하는 일종의 탁자형 터치스크린이다. 본 논문에서는 이러한 환경에서 사용자 독립적인 온라인 필기 숫자를 인식하는 연구를 수행하였다. 이로 인해 추후 진행될 다중 사용자의 한글, 영문, 특수기호의 인식 가능성을 확인하였다. 실험 과정은 테이블 탑 디스플레이의 표면을 통해 입력된 사용자별 손가락 궤적으로 기준점을 잡고, 각 사용자별 필기궤적에서 대표점 추출과 16-방향 체인코드변환을 수행하였다, 변환된 체인코드의 학습 및 필기 숫자 인식에 확률·통계적 모델인 은닉 마르코프 모델을 이용하였다. 실험에 사용된 데이터는 총 300개의 데이터를 사용 하였고, 학습은 10회 복제하여 총 3000개의 데이터로 수행하였다. 각 사용자별 데이터를 100 개씩 인식 실험에 사용하여 각각 93%, 94%의 정인식율을 보였다.

1. 서론

최근에 인간과 컴퓨터의 상호작용을 위한 입력장치 중 테이블 탑 디스플레이라는 멀티터치 입력 장치가 활발히 연구 되고 있다. 마이크로소프트사의 서피스 컴퓨터(Surface Computer)나 2006년 개발된 전남대학교 CT연구소의 테이블 탑 디스플레이(Table Top Display)가 대표적인 예 이다. 테이블 탑 디스플레이는 사람에게 친숙한 상호작용의 수단인 손을 인터페이스 수단으로 이용하는 일종의 탁자형 터치스크린으로, 입력 신호를 감지하는 센서표면과 출력신호를 재현하는 디스플레이로 구성된다. 기존의 탁자형 디스플레이 장치는 오직 한명의 사용자만을 지원하였으나 현재는 인식 기술의 발달로 멀티유저, 멀티터치를 지원하는 방향으로 발전되고 있다. 현재 테이블 탑 관련 응용기술들은 가장 흔한 지도검색, 가상키보드 및 사진 편집과 같은 응용에서 점차 벗어나고 있다. 가장 최근의 연구로서 뉴욕대는 멀티-터치 기술을 이용해 다양한 상호작용을 선보였는데 이는 테이블 탑 디스플레이의 발전 가능성과 그 중요성을 일깨워준 계기가 되었다[1].

전남대학교 CT연구소에서 개발한 테이블 탑 디스플레이 기술은 10명의 유저(손가락)가 동시에 명령을 내릴 수 있다. 현재는 사용자 독립적인 단일-터치 인식과 영역제한을 통한 멀티-터치 인식 방법은 가능하지만, 사용자간의 영역을 제한하지 않는 방법은 연구 단계에 있다.



<그림 1> 테이블 탑 디스플레이

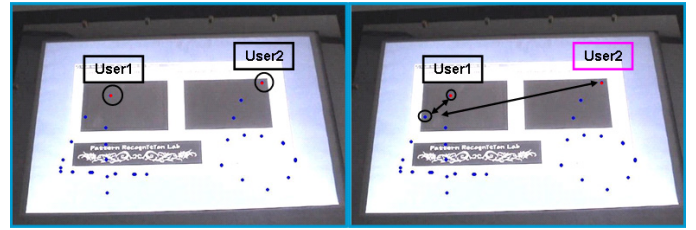
본 연구는 전남대학교 CT연구소에서 개발한 테이블 탑 디스플레이 환경에서 사용자 독립적인 동시 입력을 받아들여 각각의 사용자들이 입력한 필기체 숫자를 인식하는 시스템을 개발하는 것을 목적으로, 인식에는 확률·통계적 모델인 HMM(Hidden Markov Model)을 사용하였다. 본 연구로 인해 기존의 테블릿 PC(Tablet Personal Computer)나 PDA(Personal Digital Assiatant)등에서 사용한 단일 사용자의 필기체 문자 인식 기술에서 한층 더 발전된 다중 사용자의 온라인 필기체 문자 인식 기술 개발에 더욱 박차를 가하게 될 것이다.

2. 시스템의 구성 및 개발

2-1 시스템의 구성

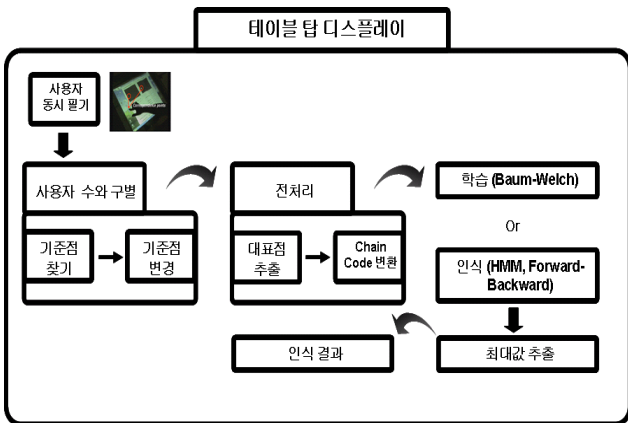
테이블 탑 디스플레이 환경에서 사용자 독립적인 온라인 필기 숫자 인식 시스템의 전체적인 구성은 <그림 2>와 같다. 우선, 다수의 필기자가 동시에 테이블 탑 디스플레이

레이에 필기를 하면 사용자 수와 사용자를 구분한다. 사용자별로 입력되는 손가락 궤적을 좌표값으로 받아 전처리 과정을 수행하게 된다. 전처리 과정을 통해 16방향의 체인코드로 변환되고, 변환된 체인코드를 학습을 통해 생성된 필기 숫자 10개(0~9)의 은닉 마르코프 모델에 입력하여 확률을 계산하고, 가장 높은 확률의 모델을 결과로 출력한다. 전처리 과정은 대표점 추출과 16방향 체인코드를 사용한다. HMM의 학습과정은 필기수식의 체인코드로 은닉 마르코프 모델을 구성하는 과정으로 Baum-Welch 알고리즘을 사용한다. HMM의 인식과정은 체인코드가 어느 은닉 마르코프 모델에 더 가까운가를 확률로써 판단하는 과정으로 Forward-Backward 알고리즘을 사용한다.



<그림 3> 사용자 기준점 정의 <그림 4> 가까운 기준점 찾기

이때 각 사용자별로 정의된 기준점은 입력이 들어올 때마다 적절하게 변화하여야 한다(그림 5). 이유는 <그림 7>처럼 필기도중 다른 사용자의 기준점에 가까워지는 경우 다른 사용자의 입력으로 오인될 수 있기 때문이다. 기준점의 변화에 사용한 수식은 다음과 같다.



<그림 2> 시스템 구성

$$S_x(i) = \frac{S_x(i-1) + I_x}{2}$$

$$S_y(i) = \frac{S_y(i-1) + I_y}{2}$$

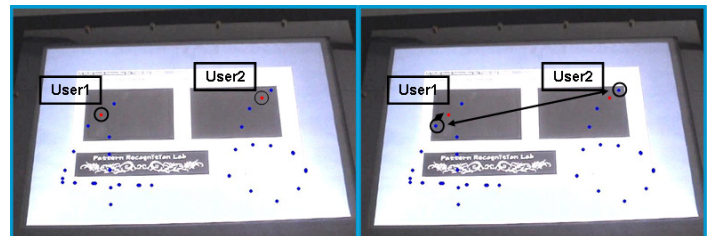
$S_x(i)$: 새로운 기준 X좌표
 $S_y(i)$: 새로운 기준 Y좌표
 $S_x(i-1)$: 현재 기준 X좌표
 $S_y(i-1)$: 현재 기준 Y좌표
 I_x : 현재 입력된 X좌표
 I_y : 현재 입력된 Y좌표

<표 1> 기준 좌표점의 변경 수식

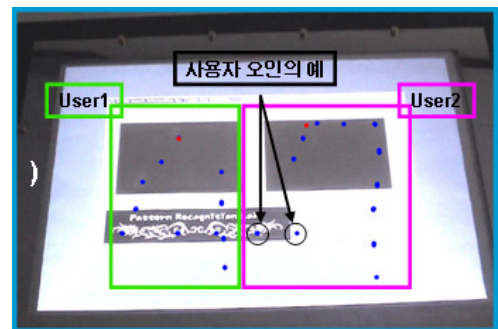
이 과정이 끝나면 다음 추출된 픽셀의 좌표와 수정된 기준점을 토대로 위의 과정을 반복하게 된다(그림3~6).

2-2 입력된 좌표의 사용자별 구분

다중 사용자가 테이블 탑 디스플레이에 동시에 숫자를 쓰게 되면, 먼저 동시에 필기를 하는 사용자의 수(동시에 터치된 손가락의 숫자)를 체크하여 각 사용자별로 최초로 터치된 좌표를 기준점(그림 3)으로 정하게 된다. 그 후 순차적으로 입력되는 좌표들에 대해서는 사용자별 기준점과의 유클리디안 거리(Eucladian Distance)를 계산하여 가장 가까운 기준점을 찾게 된다(그림 4). 그 후 가장 가까운 기준점에 정의된 사용자의 필기로 보고, 해당 좌표를 저장하게 된다. 기준점과의 거리계산이 필요한 이유는 테이블 탑 디스플레이의 특성상 모든 필기가 동시에 이루어진다면 어느 지점의 어느 좌표가 어느 사용자에 의해 필기가 되는지 쉽게 알 수 있으나, 실제적으로는 모든 사용자의 필기가 동시에 이루어지지 않기 때문에 필기 시 추출되는 좌표가 어느 사용자의 필기에 의해 추출되었는지 알 수 없기 때문이다. <그림 3>, <그림 4>는 2명의 사용자가 동시에 필기 숫자를 입력하게 되는 경우를 가정한 것이다.



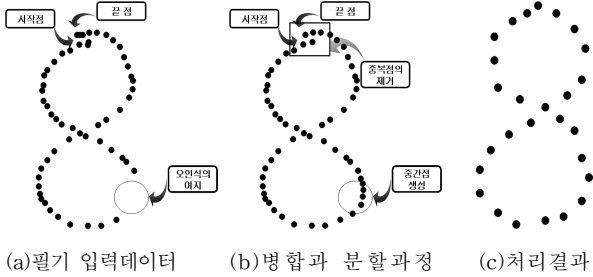
<그림 5> 기준점 변경 <그림 6> 새로운 기준점과의 비교



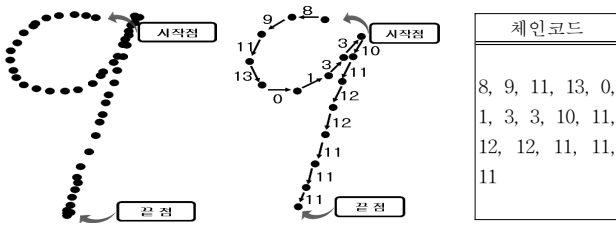
<그림 7> 기준점 불변 시 오류의 예

2-3 특징추출 및 학습 과정

특징 추출은 특정 글자들에 대한 데이터를 잡음제거 (Noise Reduction), 대표점 추출 등의 전처리 과정을 거쳐, 문자 모델들에 대한 각각의 특징을 16-방향이나 32-방향 체인코드(Chain Code)로 표현해 내게 된다(그림 8,9)[3].



<그림 8> 대표점 추출 및 전처리 과정



<그림 9> 체인코드 변환 과정

은닉 마르코프 모델을 이용한 방법의 학습은 은닉 마르코프 모델의 파라미터를 추정하는 것으로 정의할 수 있다. 이 문제는 은닉 마르코프 모델을 이용하는 데 있어 가장 어렵고 중요한 문제인데, 파라미터들은 ML(Maximum Likelihood)기준으로 학습된다. 즉, 체인코드를 가장 잘 설명하는 파라미터 λ 을 찾는 과정이다. 이 문제는 EM알고리즘으로 구현된 “Baum-Welch 재추정 알고리즘”이라고 알려진 반복 과정으로 해결한다.[패터인식개론] 은닉 마르코프 모델의 파라미터(λ)는 a 행렬, b 행렬, Π 벡터로 구성되는데, a 행렬은 상태들 간의 전이 확률을 나타낸 행렬이고, b 행렬은 각 상태에서 관측된 심벌들의 확률 분포이며, Π 는 각 상태들의 초기 분포를 의미한다. 이때 관측된 심벌은 체인코드를 말하는 것이다. 한편 학습 알고리즘은 <식3>,<식4>와 같으며, E-단계와 M-단계를 은닉 마르코프모델의 파라미터가 변화하지 않을 때까지 반복한다[2].

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} \zeta_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i,j)}$$

$$\bar{b}_j(V_k) = \frac{\sum_{t=1}^T \gamma_t(j)}{\sum_{t=1}^T \gamma_t(j)} \quad s.t. x_t = v_k$$

$$\bar{\pi}_i = \gamma_1(i)$$

<표 2> E - 단계

E-단계에서 $\zeta_t(i,j)$ 는 은닉 마르코프 모델과 사용자의 입력이 전처리를 통해 변환된 체인코드가 주어졌을 때, 시간 t에서의 상태 i와 시간 t+1에서의 상태 j의 확률이다. 이때 시간 t에서의 상태 i의 확률 $\gamma_t(i)$ 는 모든 j에 대해서 $\zeta_t(i,j)$ 를 더하는 것으로 구할 수 있다. 또한 계산된 $\gamma_t(i)$ 를 모두 합하여 상태 i로부터 일어나는 전이 횟수를 추정할 수 있다. 마지막으로 모든 시간에 대해서 $\zeta_t(i,j)$ 를 더함으로써 상태 i에서 상태 j로의 예상 전이 횟수를 구할 수 있다[2].

1. $\alpha_t(i) = P(O_{t=1}, \dots, O_{t=T} | a_T = S_i, \lambda), \text{ for } \forall t, i$
2. $\beta_t(i) = P(O_{t=T}, \dots, O_{t=1} | a_t = S_i, \lambda), \text{ for } \forall t, i$
3. $\gamma_t(i, j) = P[O_{t-1} = S_i \cdot a_t = S_j]$
 $= \alpha_{t-1} a_{ij} b_{ij}(O_t) \beta_t(j)$
 $(1 \leq t \leq T)$

<표 3> M - 단계

M-단계에서는 전향 변수 $\alpha_t(i)$ 와 후향 변수 $\beta_t(i)$ 를 먼저 정의하게 된다. $\alpha_t(i)$ 는 Forward Algorithm을 통해 구할 수 있는데, 전향 변수는 모델 λ 가 주어졌을 때, 시간 t=1에서 시간 T까지의 각 상태(S_i)에서 전처리를 통해 생성된 체인코드가 나타날 확률을 의미한다. 한편 후향 변수 $\beta_t(i)$ 는 Backward Algorithm을 통해 구할 수 있다. Backward Algorithm은 Forward Algorithm의 역순으로 진행되는데, $\beta_t(i)$ 는 모델 λ 와 시간 T에서의 상태 S_i 가 주어졌을 때, T에서부터 시간 t=1까지 체인코드가 관측될 확률로 나타낸다[2].

2-4 인식 과정

학습과정을 통해 특정 숫자에 대해 각각의 은닉 마르코프 모델의 파라미터(λ)가 만들어지게 된다. 인식은 관측된 체인코드가 발생할 가능성이 가장 높은 숫자의 은닉 마르코프 모델을 찾는 것이다. 다시 말하면 입력된 글자의 체인코드가 각각의 숫자 모델로부터 발생할 확률을 계산하고 가장 높은 확률을 지니는 모델을 찾는 것이 인식의 과정이라는 것이다. 확률계산을 위해서 각각의 모델에 Forward Algorithm을 적용하였다. Forward Algorithm은 전향 변수 $\alpha_t(i)$ 를 구하는 과정으로 알고리즘은 <식 5>와 같다. 전향 변수 $\alpha_t(i)$ 는 모델 λ 가 주어졌을 때, 시간 t에서 관측된 체인코드 O_1, O_2, \dots, O_t 와 상태 i의 확률을 뜻한다[2].

1. Initialization

$$\alpha_t(i) = \pi_i b_{(O_t)}, 1 \leq i \leq N$$

2. Induction

$$\alpha_{t+1}(i) = \left[\sum_{j=1}^N \alpha_t(j) a_{ij} \right] b_j(O_{t+1}), 1 \leq t \leq T-1, 1 \leq j \leq N$$

3. Termination

$$P(O|\lambda) = \sum_{i=1}^N \alpha_T(i)$$

<표 4> Forward Algorithm

2-5 실험 환경 및 결과

실험은 테이블 탑 디스플레이에서 이루어 졌다. 은닉 마르코프 모델의 훈련 데이터의 수집은 테이블 탑 디스플레이에서 0부터 9까지를 숫자를 30번씩 필기하여 수집하였다. 또한, 필기자에게 정상적인 필기방법으로 정체를 쓸 것을 요구하였다. 이 데이터로부터 총 300개의 체인코드를 생성하고, 10회 복제하여 총 3000개의 데이터로 만든 후, 학습을 수행하였다. 이 때 은닉 마르코프 모델에서의 상태(state)는 10개로 하였고, 학습에 필요한 상태(state)는 랜덤(random)으로 할당하였다. 체인코드는 16-방향 체인코드를 사용하였다. 인식 실험은 2명의 사용자가 동시에 숫자별로 각각 10회씩 테이블 탑 디스플레이에 필기하여, 사용자별로 정인식율을 계산하였다. 실험 결과는 <표 5>와 같이 93%, 94%의 정인식율을 보였다. 오인식의 경우는 '2','4','5'와 같이 다른 사용자의 필기에 영향을 줄수 있는 숫자에서 나타났다. 이는 숫자 '2','4','5'가 가지는 x좌표상의 이동이 많다는 특성과 가로방향의 획에 대해서 사용자별 기준점을 오인식하는 것으로 파악된다.

숫자	사용자 1		사용자 2	
	정인식	오인식	정인식	오인식
0	10	0	10	0
1	10	0	10	0
2	8	2	7	3
3	10	0	10	0
4	7	3	8	2
5	8	2	9	1
6	10	0	10	0
7	10	0	10	0
8	10	0	10	0
9	10	0	10	0
전체	93	7	94	6
	정인식율 : 93%		정인식율 : 94%	

<표 5> 실험 결과

3. 결론

본 연구는 테이블 탑 디스플레이 환경에서 다중 사용자

가 동시에 숫자를 필기하였을 때, 각각의 사용자들의 필기를 구별하여, 실시간으로 개별 사용자가 필기한 숫자를 인식하는 시스템을 구현하였다. 그 결과 성공적으로 숫자를 인식하였다. 이는 추후 진행 될 테이블 탑 디스플레이 환경에서 한글·영어·특수기호, 나아가 문장의 인식까지도 얼마든지 가능함을 의미하는 것이다.

그러나 다중 사용자들이 사각의 테이블 탑 디스플레이에서 동일선상에서 필기를 수행하지 않을 경우, 사용자 분류는 가능하나, 체인코드를 생성하는 데 있어 문제가 발생한다. 이는 인식에 있어서 치명적인 문제이다. 따라서 다중 사용자들이 서로 마주보면서 필기를 수행하는 하는 것과 같이 동일한 선상에 위치하지 않고 필기를 할 경우에 대한 연구가 추가적으로 필요할 것이다. 또한, 테이블-탑 디스플레이 기술은 인간과 컴퓨터의 자연스러운 의사소통을 위해서 반드시 해결해야 할 문제이며, 향후 지속적인 연구 개발이 필요한 분야이다. 인간이 도구에서 자유로워질 때까지, 더 나아가 네 가지 원소(인간, 컴퓨터, 물리적 객체, 투영된 객체)의 정서적인 교류가 가능해질 때까지 테이블-탑 디스플레이 분야의 연구는 꾸준히 진전되어야 할 것이다.

Acknowledgement

본 연구는 문화관광부 및 한국문화콘텐츠진흥원이 주관하는 전남대학교 문화 콘텐츠 기술 연구소 육성사업의 연구비 지원에 의해 수행되었음

참고문헌

[1] 김송국, 이철우, "멀터치를 위한 테이블-탑 디스플레이 기술 동향", 한국 콘텐츠학회 논문집 제7권 제2호, 2007, p84~91.

[2] 김의철, 김지웅, 김수형, "테이블 탑 디스플레이 기반의 온라인 필기 숫자 인식", 한국정보처리학회 추계학술발표대회 논문집 제 14권 제 2호, pp. 9~12, 2007.

[3] 김지웅, 김의철, 김수형, "테이블-탑 디스플레이 기반의 온라인 필기 문자인식을 위한 전처리 기법", 한국정보처리학회 추계학술발표대회 논문집 제 14권 제 2호, pp. 171~174, 2007.

[4] 신봉기, "HMM theory and Applications", 2003컴퓨터 비전 및 패턴인식 연구회 춘계 워크샵 튜토리얼, 2003.

[5] 신봉기, 김진형, "은닉 마르코프 모델을 이용한 온라인 한글 인식", 한국통신학회 워크샵 제1권, 1993, p189~194.

[6] 한학용, {패턴인식개론}, 한빛미디어, 2005, p468~499.

[7] Lawrence R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", IEEE Proc., 1989, p.267~295.

[8] Richard O. Duda, Peter E. Hart, David G. Stork, {Pattern Recognition 2nd ed.}, in 유현중(ed.) ITC, 2006, p148~160.