

필기영상의 동적 정보 추출 및 인식을 위한 통계적 모형

신봉기

부경대학교 전자컴퓨터정보통신공학부

bkshin@pknu.or.kr

An Accurate Stochastic Model for the Pen Trajectory-Based OCR

Bong-Kee Sin

Computer Multimedia Division, Pukyong National University

요 약

온라인 필기 인식기의 필기 모델을 응용하여 오프라인 한글 필기의 필기 궤적을 추적하고 인식하는 방법을 제안한다. 사용한 온라인 모델은 HMM의 망으로 구성된 조합형 한글 필기 모델 BongNet이다. 그리고 시계열 신호의 길이에 대한 모델이 전혀 없는 표준 HMM 대신 동적인 연속 출력 nonstationary HMM을 이용한 방법을 기술하였다. 획 추적 계산 과정에는 프레임 동기 알고리즘을 적용한다. HMM의 각 상태는 가능한 필기 궤적상의 위치에 대한 정보를 기록한다. 매 시각마다 최종 상태의 후보 중에서 모든 획을 완전히 지나는 경로가 있는지를 조사한다. 본 방법은 문자영상에서 온라인 시계열 코드를 만들어 가는 과정이며 코드와 동시에 인식결과를 출력한다.

1. 서 론

문자인식 연구 분야에서 오프라인 필기 인식 방법에 대한 연구는 오랜 역사를 가지고 있다. 하지만 인식 성능은 아직도 사람들의 기대 수준에 크게 못 미치며 제대로 된 체계적인 방법론조차 찾아보기 힘든 실정이다. 반면 80년대 말에 붐이 일기 시작한 온라인 필기 인식 분야에서는 상당히 우수한 인식기와 함께 수많은 연구 논문들이 쏟아져 나왔다. 온라인 인식 분야의 견인차 역할을 한 것은 통계적 모형의 하나인 HMM(Hidden Markov Model)이었다[1]. HMM을 이용한 온라인 필기 인식은 이론과 실제적인 면에서 매우 성공적이었기 때문에 온라인 인식 방법을 오프라인 인식에 활용하려는 연구가 자연스럽게 나타나게 되었다. 사람들의 생각은 필기 영상에서 필촉의 이동 경로(궤적)를 추출하고 그것을 온라인 정보로 바꾸어 보자는 것이었다. 그래서 D. Doerman 등은 필기 궤적 복구(pen trajectory recovery) 문제에 매달렸다[2]. 그들의 아이디어는 고해상도 필기 영상에서 임크가 묻는 방식, 농담, 빠침 등 저급 영상정보를 자세히 분석하면 궤적을 추적할 수 있다는 것이었다. 하지만 그 정보를 온라인 필기로 변환하여 실제로 인식기와 연결하거나 두 시스템을 유기적으로 통합하는 방법에 대한 연구는 아직까지 본 저자의 연구 외에는 전혀 발견할 수 없었다. 본 논문에서는 이전 연구의 소개하고 이를 확장한 모델을 제안한다.

2. 2D HMM 필기 모델

시계열 데이터와 달리 카메라/스캐너 영상과 같이 시간을 배제한 순수한 공간 패턴에는 어떤 순서도 없기 때문에 causal 관계를 도입하기가 어렵다. 하지만 최근 causality를 도입하고 그리하여 계산량을 줄일 수 있는

유사한 연구 사례가 나오고 있다[3]. 주요 아이디어를 보면 격자 과정을 구성하기 위하여 다음 두 가지 가정을 하였다. (1) 첫째, 노드간 전이를 제한하여 2차(order) 이웃 체계를 쓰고, (2) 둘째, 2D 공간의 경우 위에서 아래로, 왼쪽에서 오른쪽 방향의 causality가 있다는 것이다.

2D HMM의 노드 연결 구조는 일반적인 ergodic 모델과 같이 복잡도가 높은 구조(topology)보다 영상의 2D 평면 구조를 내포하는 그물 메쉬형 구조가 직관적으로 더 적절하다. 뿐만 아니라 2D 패턴의 모델로써 충분히 일반성이 있으며 영상의 다양한 부분적 변형을 쉽게 표현할 수 있는 구조이다. 즉 2D에서 causality는

$$r_{ij}^h = p(q_u = \lambda q_{u-v} = h), \quad h \in \{j-N-1, j-N, j-N+1\}$$

$$s_{ij}^i = p(q_u = \lambda q_{u-1} = i), \quad i \in \{j-N-1, j-1, j+N-1, j\}$$

와 같이 표현된다. 여기서 r_{ij}^h 는 노드 h 에서 노드 j 로의 수직방향 causality를, s_{ij}^i 는 노드 i 에서 노드 j 로의 수평방향 상태 천이 causality를 나타낸다.

노드 j 의 위에서 노드 j 로 천이 가능한 입력 노드의 집합을 μ_j , 왼쪽에서 천이해 들어올 수 있는 노드의 집합을 λ_j , 그리고 노드 j 에서 천이해 갈 수 있는 오른쪽 및 아래쪽 노드의 집합을 각각 ξ_j , θ_j 라고 하자. 그러면

$$\sum_{k \in \xi_j} s_{jk}^j = 1, \quad s_{ij}^i \geq 0 \quad \text{및} \quad \sum_{k \in \theta_j} r_{jk}^j = 1, \quad r_{ij}^i \geq 0$$

과 같은 확률 조건을 만족해야 한다. ξ_j 와 θ_j 는 노드 j 에서 본 이웃, 즉 천이가 가능한 이웃이다.

이와 같은 모형은 2D 영상 패턴을 완전한 2차원 모델로 표현한다는 점에서 의의가 있으나 현재 고립 글자(isolated character) 인식에만 활용하는 방법만 개발되어 있어서 자소 합성에 의한 한글 인식에는 바로 적용하기 어렵다. 자세한 내용은 [3]을 참조하기 바란다.

3. 동적 정보 추출

동적 정보 또는 필기 궤적 추출(pen trajectory recovery, PTR)은 D. Doerman 등이 가장 체계적으로 연구를 하였다[2]. 하지만 그 이후로 후속 연구가 거의 없었고 오프라인 문자 인식과 연계된 연구는 본 저자가 아는 한 더더욱 없었다. 어쨌든 PTR 방법의 성과가 좋다면 오프라인 인식 문제의 돌파구를 마련할 수 있다는 점에서 PTR 연구는 아직도 가치가 있다고 생각된다. 실제로 상당히 알아보기 힘든 달필, 난필, 악필의 경우 사람도 한 자씩 더듬더듬 필기 궤적을 따라 가면서 읽는다는 사실은 PTR 연구를 더욱 정당화해준다.

Doerman은 영상 분석에만 매달렸다. 필기 도구 및 종이의 특성, 필기 운동의 물리적 제약 등에 따른 획 영상의 특징을 분석하여 영어 필기체 영상에서 궤적을 추적하였다. 영어 필기체의 경우 상하 진동을 반복하고 오른쪽으로 전진하며 획이 길게 이어가는 경향이 있기 때문에 상당한 결과를 얻을 수 있었다. 하지만 한글의 경우 개념적으로 사각형의 구역 안에서 획이 상하 좌우로 배치가 되므로 어떤 방향으로 이어졌는지 추정하기가 쉽지 않다.



그림 1. 필순을 추정하는데 힌트가 있는 패턴의 예.

문자 인식 또는 문자 인식 방법과 연계를 시도한 것으로는 저자의 과거 연구가 있다. 온라인 필기 모델을 필기 궤적 진행 방향을 추정하는데 가설 생성에 사용하고 영상에서 얻은 데이터로 검증 및 최적 후보를 선택하는 방식을 제안하였다. 각 (시간적) 단계에 모델의 각 상태에 (공간적) 위치에 대한 후보 (t, x, i) 를 생성하는데, 동적 프로그래밍 방법에 의하여 최적의 후보를 계산하는 방식을 제안하였다.

4. 온라인 필기 모델 BongNet

한국 사람 절대 다수는 어린 시절 배운 대로 초성, 중성, 종성의 순서로 글자를 쓰며, 또 각 자모도 일정한 필순에 따라 쓴다. BongNet은 이러한 필순을 자소 HMM의 망 구조에 담은 온라인 한글 음절 필기 모델이다 [1]. BongNet의 각 요소 HMM은 자소 별로 분리한 온라인 데이터를 모아 훈련하였다. BongNet에 특이한 것은 각 자소의 클러스터링 하고 연결획(ligature)을 독립적인 HMM으로 모델링 한다는 점이다. 바로 이 점 때문에 무 제약 한글 음절 필기를 인식할 수 있으며, 또 완전한 필획의 궤적을 확률적으로 표현하므로 오프라인 필기에서 필기 궤적을 추정하는데 활용할 수 있다.

BongNet의 시작 노드에서 끝 노드까지 연결하는 하나의 완전한 경로는 한 글자의 필기 궤적에 대응한다. BongNet에서 최적 경로란 결국 인식결과를 산출하는 최

적 경로이다. 본 연구에서는 오프라인 필기 궤적을 추적하는데 가장 유용한 정보, 많은 도움을 준 경로라고 할 수 있다. 오프라인 필기와 온라인 모델의 통합이란 필기 영상에 대하여 최적 길이의 최적 경로를 찾는 것이며 이는 곧 필기 인식에 해당한다.

5. Nonstationary HMM

HMM은 다양한 변이에 대한 모델링 능력이 우수하여 여러 가지 문제에 활용되고 있지만, 두 가지 기본 가정 때문에 모델 자체와 모델의 동작이 정확하지 않다는 태생적 아킬레스 건도 갖고 있다. 우선 HMM의 마르코프 가정을 보면 체인 프로세스에서 한 단계 이전의 과거 이력을 완전히 무시하기 때문에 현 상태의 반복 또는 지속 여부를 전혀 모른다는 것이다. 따라서 상태의 반복은 단순히 독립이며 따라서 상태 지속 확률은 아래와 같은 감쇄형(exponential decay) 지수함수 분포를 갖는다.

$$p(d) = a_{ii}^{d-1}(1 - a_{ii})$$

표준 HMM에서 경로 확률의 최대 값은 경로 상의 특징 천이만 최대한 반복하고 경로상의 나머지 천이는 최소한으로 하는 것이다. 이는 원래 모델 최적화 과정에서 추구한 것과는 거리가 먼 것이다. 따라서 모델의 동작 또는 모델 자체가 정확하지 않다는 것을 나타낸다.

이와 같은 문제를 해결하기 위해 제안된 상태 지속형 HMM에서는 $p(d)$ 를 별도의 분포 함수 또는 히스토그램으로 표현한다[5]. 그리고 매 시각마다 상태 천이하는 것이 아니라 일정 시간 d 후에 천이한다. 이 때문에 상태 지속 모델을 절반의 마르코프 모델이라 부르기도 한다. 이와 달리 신봉기[4]의 nonstationary HMM은 정식의 마르코프 모델처럼 동작하되 반복되는 같은 천이라도 천이 시점에 따라 천이 확률이 바뀌는 형태이다.

Nonstationary HMM(NSHMM)은 마르코프 모델의 일종으로 상태 천이 확률이 고정 상수가 아니라 지속 시간에 따라 변하는 시변형 HMM이다. 이산 출력 분포 모델과 연속 분포 모델이 제안되었으며[3] 본 연구에서는 연속 분포 모델을 활용하는 방법을 위주로 하여 비교 설명한다.

NSHMM의 출력 분포 함수는 다음과 같이 가우스 함수의 가중 결합한 가우스 결합 분포(Gaussian mixture density)를 사용한다.

$$b_j(\mathbf{x}) = \sum_{k=1}^M c_{jk} b_{jk}(\mathbf{x}_j) \\ = \sum_{k=1}^M c_{jk} N(\mathbf{x}, \mu_{jk}, \Sigma_{jk})$$

물론 조건 $\sum_{k=1}^M c_{jk} = 1$ 을 만족해야 한다. 그러면 NSHMM의 상태(Q)-출력(X) 결합 확률은

$$f(X, Q|\lambda) = \prod_{t=1}^T \prod_{i=1}^{r_{q_t}} a_{q_t, i_{t-1}}(\tau) b_{q_t}(\mathbf{x}_t)$$

이 된다. Q에서 각 상태의 가우스 함수를 하나씩 묶어 길이 T의 가우스 함수열을 $K = (k_1, k_2, \dots, k_T) \in \Omega^T$,

