

# 이미지 기반 확률모델을 이용한 문자검출

김선규\*, 신봉기\*\*

\*부경대학교 정보통신공학과

\*\*부경대학교 컴퓨터멀티미디어공학과

## Character spotting using image-based stochastic models

Seon-Kyu Kim\*, Bong-Kee Sin\*\*

\*Pukyong National University, Dept. of Telematics engineering

\*\* Pukyong National University, Division of Computer Multimedia Engineering

### 요 약

본 논문에서는 의사 2 차원 은닉 마르코프 모델의 구조로 생성한 마르코프 체인형 확률모형에 의한 인쇄체 문자 이미지의 모델링에 대해 논한다. 이미지 데이터에서 바로 모델을 실시간 생성하며 문자 인식 및 검출에 응용할 수 있다. 실험에 의하면, 이 방법을 통해 특정 낱말이 포함된 문장에서 숫자를 인식, 한글을 검출할 수 있음을 확인하였다.

#### 1. 서 론

은닉 마르코프 모델(Hidden Markov Model)은 순차적인 신호를 모델링 하는 방법으로, 음성인식 및 온라인 필기 인식 분야에 적용되고 있다. 본 논문에서 적용하고자 하는 분야는 문서이미지의 텍스트 영역 상에서 특정문자 또는 낱말을 자동으로 검출하는 것이다. 문서 이미지 상에서 문자들은 2 차원 구조이므로, 2 차원 패턴을 모델링 할 수 있는 HMM 이 필요하다. 그러나 몇몇 연구에 따르면 완전한 2 차원 HMM 은 계산량이 너무 많아 인식 알고리즘을 비효율적으로 만드는 단점이 있다.[1]. 이러한 문제점 때문에 여러 가지 모델 네트워크의 연결을 줄이는 방법이 제안되었는데, 그 중 잘 알려진 방법이 의사 2 차원 HMM[2]과 Markov random field 이다[3].

본 논문에서는 의사 2 차원 HMM 구조의 Markov-like 확률모형으로 문자 이미지를 모델링하는 방법에 대해 논한다. 2 장에서 문자 모형을 생성하는 방법을, 3 장에서는 문자 검출을 위한 세 가지 기본참조모형을, 4 장에서는 연속으로 수행결과를 얻어내기 위해 수정된 2 단계 DP 알고리즘을, 5 장에서는 본 논문에서 제안하는 방법을 이용한 몇 가지 실험 및 결과를 제시한다.

#### 2. 이미지 기반 모델

참조모형 제작은 다음 세가지 단계로 이루어진다.

(1) 이미지 샘플 평균: 첫 단계에서는 픽셀 대 픽셀 단위로 이미지의 샘플평균을 구한다. 여러 샘플을 가지고 한 개의 모델을 만들 때, 모형의  $(i,j)$  위치의 샘플 평균은 다음과 같이 구할 수 있다.

$$n_{ij} = \frac{N_{ij}}{N}$$

여기서  $N$  은 주어진 이미지 샘플의 개수이고  $N_{ij}$  는  $(i,j)$  위치가 검은 픽셀인 이미지 샘플의 개수이다.  $n_{ij}$  는  $(i,j)$  위치에서 검은 픽셀이 나타날 확률이 된다. [그림 1]은 숫자 모형에 대한 이미지 샘플 평균 예제를 나타낸 것이다.

[그림 1] 숫자 모형에 대한 이미지 샘플 평균

(2) Markov-like 모델로 변환: 두 번째 단계에서는 이미지 모형을 통계적 모형으로 변환한다. 우선 사각형 비트맵 상에서 모든 픽셀들의 세로 열에 대해 하나의 슈퍼 노드를 만들고 각 노드를 확률 1 의 전진 천이확률을 가지는 마르코프 체인으로 연결한다. 슈퍼 노드는 이미지 픽셀들의 수직 배열을 모델링한다. 모델의 수직 구조는 슈퍼 노드와 유사하게 서브 노드들의 연결로 구성하면 되는데, 한 수직 프레임  $x_j$  상에서 픽셀  $x_{ij}$  는 이미 (1)에서 계산된 통계치  $n_{ij}$  를 가지는 서브 노드로 구성된다. 그리고 각 서브 노드들은 확률 1 의 전진 천이확률로 연결되어 있다.

(3) 상태 병합에 의한 정보중복성 축소: 마지막 단계로 모델링의 정확성을 해치지 않는 범위에서 글자 모형이 가지는 정보중복성을 제거한다. 이러한 정보중복성 감소는 두 단계로 이루어진다. 첫 단계로 유사한 노드들로 구성된 수직 열을 서로 병합한다. 수직 열간의 유사도는 다음과 같은 거리함수로 측정한다.

$$d(n_i - n_j) = \sum_{k=1}^S |n_{ik} - n_{jk}|^\alpha, \alpha \in R$$

측정된 거리함수의 값이 주어진 기준치 보다 적으면, 노드 열간의 병합이 이루어지고, 해당 노드들은 다음 식에서처럼 각 픽셀 확률의 가중치 평균을 가지는 새로운 노드로 치환된다.

$$\bar{n}_{ik} = w_i n_{ik} + w_j n_{jk}$$

정보중복성을 감소시키는 두 번째 단계에는 서브노드 간의 병합을 수행한다. 각 노드의 밝기분포가 유사한 서브노드 간에 병합을 수행한다.

두개의 노드를 병합할 때, 노드간의 연결도 조건에 맞게 변해야 한다. 마르코프 모델의 전이 확률과 비교할 때, 이미지 기반의 모형은 DP정합의 삽입확률과 치환확률로 변환할 수 있다. 병합을 하기 전에, 치환확률은 1이고 삽입확률은 0이다.

### 3. 검출 모델

#### 3.1 Key 모형

Key 모형을 만들 때, 여러 가지 패턴으로 나타나는 글자를 수용할 수 있도록 모든 문자마다 하나의 모형을 할당할 필요가 있다. 숫자의 경우 10개의 모형이면 충분하다. 그러나 한글과 같은 동양의 글자들의 경우에는 문자의 개수가 너무 많으므로 모델을 동적으로 구성하는 방법이 필요하다.

#### 3.2 Filler 모형

Filler 모형은 Key 문자를 제외한 다른 문자와 문자가 아닌 기호 등과 정합하기 위한 모형이다. Filler 모형을 만들기 위해 모든 이미지 샘플들의 평균을 취하고 앞장에서 설명한 방법대로 모형을 만든다. [그림 2]는 숫자 인식을 위한 filler 모형의 비트맵 이미지이다.



[그림 2] Filler 모형의 비트맵 이미지

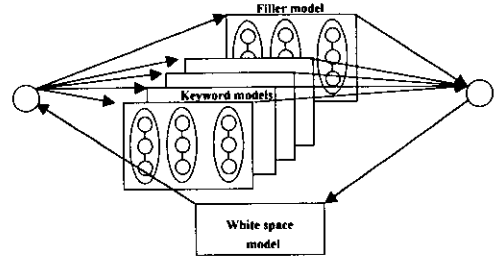
한글과 같이 글자의 형태로 구별이 힘든 문자 부류에 대해서는, 둘 이상의 filler 모델을 적용할 수 있다.

#### 3.3 White space 모형

Key 모형이나 filler 모형은 입력된 이미지상에서 문자영역을 나타내는데 사용되고, 나머지 부분은 공백으로 취급한다. 공백은 문자의 부분이 아니므로, 공백과 정합할 수 있는 특별한 하나의 모형이 필요하다. White space 모형은 매우 간단하고 노드의 수도 다른 모형에 비해 아주 적다.

### 3.4 검출 네트워크 모델

문자 검출을 위해 여기서는 네트워크 기반 순환구조의 전이모델을 설계했다. [그림 3]의 검출 네트워크 모델은 길이와 순서가 정해지지 않은 Key 패턴 및 다른 패턴까지도 모델링 할 수 있다.



[그림 3] 숫자에 대한 검출 네트워크

입력이미지의 프레임들은 왼쪽의 dummy 노드에서 시작해서 모든 가능한 경로를 차례로 탐색하게 된다. 가장 좋은 길이의 가장 좋은 결과를 탐색하는 것은 다음 장에서 논할 것이다.

### 4. Key 문자 탐색 알고리즘

탐색이 계속되는 동안 연속으로 결과를 얻기 위해 개선한 DP 탐색 알고리즘을 적용하였다. 탐색 알고리즘에 적용할 입력으로 SXT 블록이미지  $X$ 가 있다고 하자. 탐색 프로세스는 이미지의 왼쪽에서 오른쪽으로 한 수직 프레임  $X_i$  단위로 스캐닝하게 된다.

$$X = \{x_{st}\}_{S \times T} = X_{1,T} = \{X_t : t = 1, \dots, T\}$$

그리고,  $K$  개의 모형  $r^k$  을 가정하자.

$$r^k = (r_1^k, \dots, r_N^k), k = 1, \dots, K$$

여기서  $r_i^k$  는 어떤 문자에 대한  $i$  번째 수직 모형을 나타낸다.

DP 탐색의 주 목적은  $X$  와  $r^k$  간의 정합 확률을 계산하는 것이다. 여기서 다음과 같은 최소 정합 확률을 정의할 수 있다.

$$D(X_{1,T}, r^k) = \min_w \sum_{t=1}^T d(X_t, r^k(w(t)))$$

$w: \{1, \dots, T\} \rightarrow \{1, \dots, N\}$  은 입력 프레임  $X_t$  을 노드  $r(w(t))=r(j)$ ,  $j \in \{1, \dots, N\}$  과 사상하는 warping 함수이다.  $D(X, r^k)$  은 서로 다른 길이의 입력과 노드 사이의 거리를 나타낸다.

주어진  $r_j^k = r_1^k, r_2^k, \dots, r_j^k$  에 대해  $d(X_t, r_j^k)$  는 입력 프레임  $X_t$  와 노드  $r_j^k$  간의 거리를 나타낸다.

$$D(X_{1,t}, r_j^k) = \min \{ D(X_{1,t-1}, r_{j-1}^k) + d(X_t, r_j^k), D(X_{1,t-1}, r_j^k) + d(X_t, r_j^k) \}$$

위 식의 계산은  $t=1, \dots, T, j=1, \dots, N^k$  에 걸쳐 반복된다. 여기서  $N^k$  는  $k$  번째 모형의 노드 개수이다.

위의 식만으로 아직 프레임·노드간 거리는 알 수 없는데, 다음과 같이 표현할 수 있다.

$$d(X_i, r_j^k) = \min_w \sum_S d(x_{si}, r_{w(s)}^k)$$

거리는 DTW의 거리함수처럼 수직 space-warped 거리함수로 정의한다. 여기서  $X_i = x_{1i}, x_{2i}, \dots, x_{Ni}$ ,  $r_j^k = r_{1j}^k, r_{2j}^k, \dots, r_{Mj}^k$ 이다.  $w(\cdot)$ 은 앞에서 언급된  $W(\cdot)$ 와 유사한 공간 warping 함수이다.

위의 내용을 정리해볼 때, 2 단계 DP는 다음과 같은 두개의 수식으로 요약할 수 있다.

$$D^*(b, e) = \min_k D(\mathbf{X}_{b,e}, \mathbf{r}^k)$$

$$D^* = \min_{1 \leq l \leq L} \{ D_l(e) = \min_{1 \leq b \leq e} [D^*(b, e) + D_{l-1}(b-1)] \}$$

경계 조건은 다음과 같다.

$$D_0(0) = 0$$

$$D_0(e) = D(1, e)$$

$$D_l(0) = \infty, l = 1, \dots, L$$

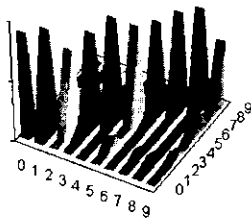
여기서  $L$ 은 인식 글자들의 최대 개수이다. 이 알고리즘은 탐색과정에서 연속적으로 결과를 얻을 수 있게 변형이 가능한다, 이것은 탐색과정에서 조건부 backtrace 기법을 이용하면 된다. 알고리즘이 수행되는 과정에서 동작 결과가 나올 때 마다, 이미 출력된 결과들은 메모리에 상주시킬 필요는 없으며, 따라서 실제 동작에서는 공간 복잡도를 줄일 수 있다.

### 5. 실험

본 논문에서 제안한 문자 검출 방법을 이용해 숫자 문자열 이미지와 한글 문자열 이미지에 대해 실험하였다. 모델을 만들 때 사용한 문자 샘플은 한 모델당 20개이며, 노이즈를 포함한 여러 가지 변화를 사용할 수 있도록 모델을 평활화 하였다. 실험에 사용된 문자는 모델을 만들 때 사용한 것과는 다른 텍스트 이미지를 사용했다.

첫 번째 실험은 숫자 문자열 인식 실험이다. 이 경우 모든 key 모형을 검출 네트워크에 포함되어 있으며, 따라서 모든 숫자를 인식할 수 있다.

실험 결과는 [그림 4]와 같다. [그림 4]는 오인식률을 보여주는 혼동행렬이다. 그림에서 볼 수 있듯이, 숫자 9의 경우 36.8%의 에러비율을 나타낸다. 전체 인식률은 총 180 샘플에 대해 88.9%의 인식률을 보이고 있다.



[그림4] 숫자 인식

두 번째 실험은 한글 문자에 대해 수행하였다. 실험은 다음과 같은 측정 방법으로 51.9%의 인식률로 측정되었고 실험결과는 [표.1]에 나타내었다.

$$\frac{H}{H + FP + FN}$$

여기서  $H$ 는 정확하게 인식한 글자의 개수를 나타내고,  $FP$ 는 부정확한 글자를 인식했을 때의 개수(false positive),  $FN$ 은 정확한 글자를 인식하지 못한 경우의 개수(false negative)이다.

한글은 모음의 형태와 종성문자의 존재여부에 따라 6가지 형식으로 구분되므로, 문자 검출 결과는 다음 표에서처럼 다시 평가할 수 있다.

[표.1] 문자 및 6형식 검출 결과

문자검출	H(%)	FP(%)	FN(%)	Remark
	51.9	48.1	0.0	
형식검출	78.9	9.9	11.3	Overall result
	70.8	0.0	29.2	1형식 (e.g. '가')
	85.7	0.0	14.3	2형식 (e.g. '간')
	100.0	0.0	0.0	3형식 (e.g. '고')
	100.0	0.0	0.0	4형식 (e.g. '곤')
	65.0	35.0	0.0	5형식 (e.g. '과')
100.0	0.0	0.0	6형식 (e.g. '관')	

본 연구는 키워드 검출에 목적을 두고 있으므로, 전체적인 성능의 향상을 위해서는 언어학적인 분석이 반영된 단어 패턴에 대한 평가가 필요할 것으로 여겨진다. 그것을 바탕으로 할 때, 본 논문에서 제안한 모델링 방법에 대한 정확한 성능 평가를 할 수 있다.

### 6. 결론

본 논문에서 제안한 문자 이미지를 모델링하는 방법은 매우 효율적이다. 실험결과에 따르면 본 논문에서 제안한 방법이 글자 인식 및 검출에 있어서 원활하게 적용될 수 있음을 알 수 있다. 추후에 이미지를 이용한 글자 모델링 방법은 다차원 비트맵 이미지 통합과 알고리즘 파라미터의 평가를 통해 더욱 개량될 것이다. 또한 실시간으로 한글의 자소 모형을 조합하여 글자의 합성 모형을 생성하는 연구도 수행될 것이다.

### 참고문헌

- [1] E. Levin and R. Pieraccini, "Dynamic planar warping for optical character recognition," in *Proc. 1992 ICASSP*, San Francisco, California, Vol. 3, pp. 149-152, 23-26 March 1992.
- [2] O. E. Agazzi and S. Kuo, "Hidden Markov model based optical character recognition in the presence of deterministic transformations," *Pattern Recognition*, Vol. 26, No. 12, pp. 1813-1826, 1993.
- [3] R. Chellapa and S. Chatterjee, "Classification of textures using Gaussian Markov random fields," *IEEE Trans. on ASSP*, Vol. 33, No. 4, pp. 959-963, Aug. 1985.