

사전기반 후처리를 이용한 모바일 폰 영상에서 와인 라벨 문자 인식

(Wine Label Character Recognition in Mobile Phone Images using a Lexicon-Driven Post-Processing)

임준식[†] 김수형^{**}
(JunSik Lim) (SooHyung Kim)

이철우^{**} 이귀상^{**}
(ChilWoo Lee) (GueeSang Lee)

양형정^{**} 이명은^{***}
(HyungJung Yang) (MyungEun Lee)

요약 본 논문에서는 모바일 폰에서 오프라인 필기체 과분할 인식의 후처리 방법에 관하여 논하였다. 제안된 방법은 조합 행렬 생성, 문자 조합 필터링, 문자 유사도 측정으로 구성된다. 조합 행렬 생성 과정은 각각의 조각의 인식 결과로부터 생성가능한 모든 조합 행렬을 계산하는 부분이며 조합 행렬을 그래프로 구성하게 된다. 문자 조합 필터링 과정은 그래프의 노드들과 단어 사전을 비교하여 불필요한

노드를 삭제하는 과정이며 문자 유사도 측정과정은 단어 사건의 각각의 단어들과 Levenshtein 거리(distance)를 계산하여 최적의 후처리 결과를 추출하게 된다. 제안된 방법의 인식률은 85.8%의 정확도를 보였다.

키워드 : Cursive Script 인식, 사전기반 후처리, 과분할

Abstract In this paper, we propose a method for the postprocessing of cursive script recognition in Wine Label Images. The proposed method mainly consists of three steps: combination matrix generation, character combination filtering, string matching. Firstly, the combination matrix generation step detects all possible combinations for a recognition result for each of the pieces. Secondly, the unnecessary information in the combination matrix is removed by comparing with bigram of word in the lexicon. Finally, string matching step decides the identity of result as a best matched word in the lexicon based on the levenshtein distance. An experimental result shows that the recognition accuracy is 85.8%.

Key words : Cursive Script Recognition, Lexicon-Driven Post-Processing, Over-Segmentation

1. 서론

하드웨어의 발전은 다양한 형태의 응용분야에 대한 사용자의 요구를 유발한다. 최근의 모바일 기기는 저전력, 고효율, 저렴한 가격의 하드웨어들이 개발됨에 따라 단순 음성 통화에서 영상 저장, 재생, 데이터 통신 및 영상 처리 기술 등을 적용한 응용분야로 사용 영역이 확대되고 있다. 이러한 하드웨어적 기술의 발전을 바탕으로 모바일 기기 사용자에게 다양한 정보를 제공하기 위한 목적으로 고해상도 카메라를 장착한 모바일 기기를 사용한 다양한 응용분야의 연구가 진행되고 있다 [1-3]. 본 논문에서는 와인 영상내의 필기문자를 인식함으로써 사용자에게 다양한 형태의 정보를 제공하기 위한 응용을 논하고자 한다. 와인 라벨은 영상 내의 필기문자는 생산지별로 다양한 언어, 형태, 글자체를 가지고 있으며 다양한 배경 및 색상으로 구성된다. 이는 자연영상 내의 필기 문자 인식을 위한 훈련 집단으로써 충분한 가치가 있다.

현재 자연영상 텍스트인식에 대하여 많은 연구가 진행 중인데, 미국 IBM Almaden 연구소에서는 카메라가 부착된 핸드폰이나 PDA 등과 같은 개인 디지털 입력장치를 통해 현장에서 획득한 이미지로부터 텍스트를 추출하고 자국의 언어로 번역하여 제공하는 연구를 수행하였고[1], Carnegie Mellon 대학의 Interactive 시스템 연구소에서는 모바일 장치에서 자연영상으로부터 추출한 텍스트를 인식하여 예제 기반 번역을 수행하는 연구

· 본 연구는 지식경제부 및 정보통신산업진흥원의 대학 IT연구센터 지원 사업의 연구결과로 수행되었음(NIPA-2010-C1090-1011-0008)

· 이 논문은 제36회 추계학술발표회에서 '과분할 및 조합문자 필터링을 이용한 와인 라벨영상에서의 Cursive Script 인식'의 제목으로 발표된 논문을 확장한 것임

[†] 학생회원 : 전남대학교 전자컴퓨터공학과
goethe0708@iip.chonnam.ac.kr

^{**} 종신회원 : 전남대학교 전자컴퓨터공학과 교수
shkim@chonnam.ac.kr
leecw@chonnam.ac.kr
gslee@chonnam.ac.kr
hhyang@chonnam.ac.kr

^{***} 정회원 : 전남대학교 전자컴퓨터공학과
melee@chonnam.ac.kr

논문접수 : 2009년 12월 29일

심사완료 : 2010년 2월 12일

Copyright©2010 한국정보과학회 : 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

정보과학회논문지 : 컴퓨팅의 실제 및 레터 제16권 제5호(2010.5)

를 진행하였다[2].

류코쿠 대학과 시가대학에서는 모바일폰 기반 카메라를 이용하여 영상에서 텍스트를 인식에 관한 연구를 수행하였고 류코쿠 대학의 NHK 과학 기술 연구소에서는 디지털 카메라로 획득된 자연 영상으로부터 일본 문자를 검출하여 영어로 변환하는 연구를 수행하였다[3]. 중국 Tshinghua 대학의 OCR연구소에서는 자연 영상으로부터 획득한 영상에서 텍스트의 위치를 찾아 인식하는 연구를 진행하였다[4].

2. 관련연구

문자 인식의 관점에서 기존 연구를 살펴보면 문자 인식의 궁극적인 목적은 오류율이 0%에 근접하도록 노력하는 것이다. 일반적인 문자 인식과정은 이진화, 문자분할, 인식, 후처리로 구성된다.

인식 과정의 성능 향상을 위하여 특징 추출과 특징 분류 두 가지 측면에서 다양한 연구가 진행 중이다. 특징추출과 특징 분류는 서로 밀접한 연관 관계를 갖고 성능 향상을 위해 적용 분야에 최적화된 방법에 관한 연구가 진행 중이다.

특징 추출에 관한 연구들은 크게 구조적 방법과 수치적 방법으로 구분할 수 있다[5]. 구조적 특징은 입력 분자의 기하학적인 모양에 기반을 두고 사람이 직관적으로 이해할 수 있는 형태학적 분석을 수행한다[6,7]. 수치적 특징은 입력 분자의 형태적 변환을 수행하여 수치적인 결과로 변환하는 과정을 수행하게 된다[5].

인식 결과에 대한 후처리는 이진화 및 분할과정에서 발생한 잡음, 인식기의 오분류 등 인식과정에서 발생한 오인식을 교정하여 성능 향상을 위하여 사용된다. 후처리 방법은 문맥적 지식의 확률적 표현에 기초한 방법, 문맥적 지식의 구조적 표현과 구조적 표현을 복합적으로 고려하는 방법으로 구분할 수 있다. 문맥적 지식의 확률적 표현에 기초한 방법은 Viterbi 알고리즘 혹은 수정된 Viterbi 알고리즘이 있으며[8], 구조적 표현에 기초한 방법은 사전에 있는 가장 유사한 단어를 탐색하는 사전 기반 비교 검색 알고리즘과 문자열 정합(String matching) 알고리즘[9], 문맥의 연결성분에 기초한 이진 N-gram 알고리즘이 있다. 복합적 방법에는 사전기반(Dictionary) Viterbi 알고리즘과 예측교정 알고리즘이 있다. 이러한 방법들은 대부분 문맥적 지식을 이용하여 후처리 작업을 수행함으로써 인식을 향상에 한계가 있으며 잡음이 많이 포함된 인식결과에 대한 적용이 어렵다.

본 논문에서는 모바일 환경에서 필기체 인식의 성능 향상을 위하여 그래프, 문자 집합(Character set), Levenshtein 거리를 이용한 복합적 후처리 방법을 제안한다.

3장의 1절에서 과분할 인식에 대하여 간략하게 설명하고 2절에서는 그래프를 이용한 인식결과 필터링 방법에 대하여 설명한다. 3.3절에는 Levenshtein 거리를 이용한 잡음 영역과 문자 영역의 분리에 대하여 설명하고 4절에서는 유사도 측정방법에 대하여 설명한다. 4장에서는 실제 모바일 폰에서의 실험 결과에 관하여 기술한다.

3. 제안 알고리즘

제안 알고리즘은 크게 과분할, 인식, 후처리로 구분할 수 있으며 그림 1과 같다. 과분할 과정에서는 입력 영상에 대하여 이진화 및 과분할을 수행하게 되며, 인식 과정에서는 분할된 조각들에 대하여 조합 영상을 생성하고 조합된 영상을 인식한다. 이때 조각의 수에 따른 처리 시간지연의 문제가 발생하므로 후처리 과정에서 조합 문자 필터링을 수행하게 된다. 최종적으로 사전에 포함된 단어와 인식 결과를 비교하여 인식 결과를 도출하게 된다.

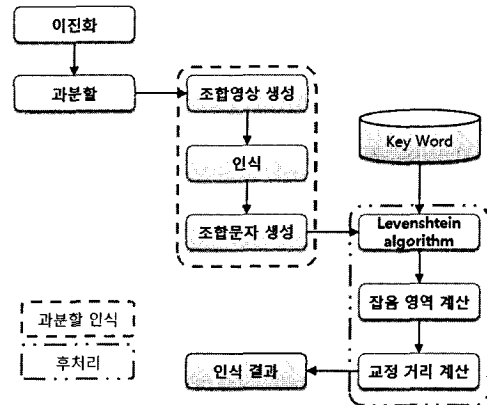


그림 1 제안 알고리즘 구성도

3.1 과분할(over-segmentation)

과분할이란 입력으로 주어진 단어 영상을 하나의 문자보다 같거나 작은 단위인 이미지 조각으로 잘게 분리하는 것으로 입력영상의 획을 분석하여 분리하게 된다. 과분할은 일반적인 문자 분할이 어려울 경우 사용되는데 그림 2와 같이 문자가 서로 겹쳐있거나 한 영상 내의 문자들간 크기가 클 경우 효과적이다. 본 논문은 문자

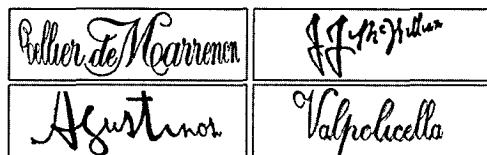


그림 2 문자단위 분할이 어려운 필기체 영상

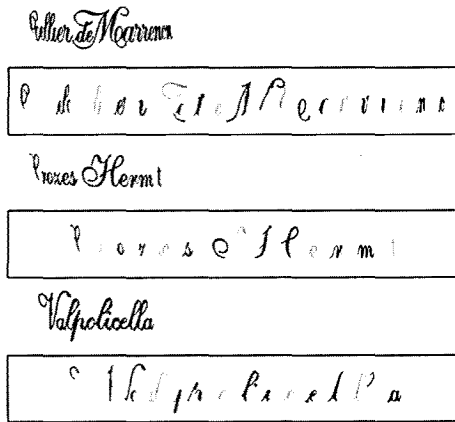


그림 3 과분할의 예

분리를 위하여 과분할 방법을 이용하여 분할을 수행하였고 그림 3은 분할 결과의 예이다.

과분할을 통한 인식방법은 분할된 순서대로 조합을 수행하여 인식결과를 추출하게 된다(그림 4). 이러한 인식결과를 인식 대상 문자와 비교하기 위해서는 조합이 가능한 문자조합을 선택하여 조합 문자를 생성하여야 한다.

여기서 조합이 가능한 문자집합이란 다음과 같다. 인식 결과를 R이라 하고 조합된 문자의 인식결과를 $r_{(a,b)}$ 로 정의하였을 때(여기서 a는 조합이 시작되는 조각, b는 조합에 사용된 조각의 수) $r_{(0,3)}$ 과 $r_{(1,1)}$ 은 조합이 불가능하다. 그 이유는 $r_{(0,3)}$ 에서 0,1,2번 조각을 이용하여 조합을 수행하였으므로 1번 조각을 사용하는 $r_{(1,1)}$ 은 조합에 사용 될 수 없다. 이러한 조건을 만족하는 문자 조합을 생성하여 후처리 과정에서 사용하게 된다. 하지만 표 1과 같이 분할과정에서 발생하는 조각의 수에 따라 조합문자집합의 수가 증가하고 이에 따른 후처리 시간의 증가를 초래한다.

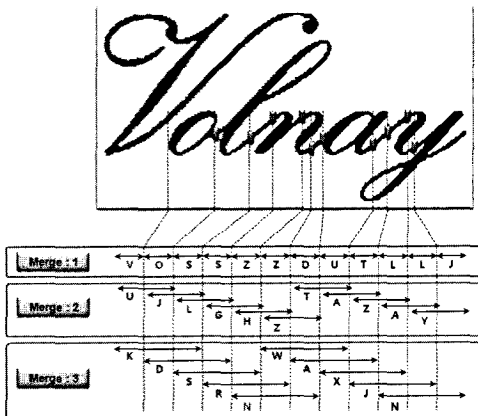


그림 4 조합 및 인식 결과

표 1 조각 수에 따른 조합 문자수의 증가

분할조각 수	조합문자 수	분할조각 수	조합문자 수
1	1	11	504
2	2	12	927
3	4	13	1,705
4	7	14	3,136
5	13	15	5,768
6	24	16	10,609
7	44	17	19,513
8	81	18	35,890
9	149	19	66,012
10	274	20	121,415

3.2 인식결과 필터링

문자 조각 수에 따른 조합 문자열의 수가 증가하는 문제를 해결하기 위하여 인식 결과를 그래프로 구성하였으며 인식 결과와 비교대상문자에 포함된 문자들을 비교하여 노드의 삭제 수행한다.

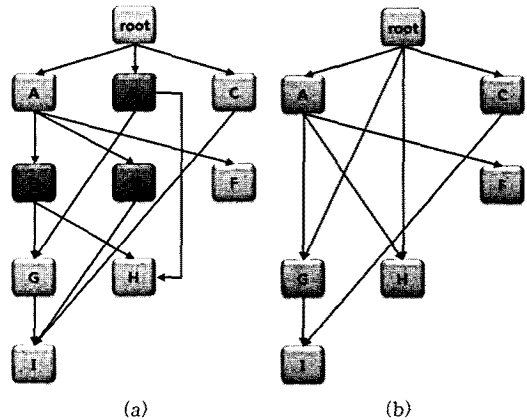


그림 5 그래프로 구성한 인식결과: (a) 그래프 구성, (b) 삭제대상 노드 제거 후

과분할 인식 결과에는 정확한 문자의 구성을 예측할 수 없으므로 불필요한 정보가 포함되게 된다. 불필요한 인식결과를 제거하기 위하여 인식결과와 제거과정을 수행한다. 비교 대상 문자에 포함된 문자열의 bigram과 비교하여 불필요한 연결들을 제거하게 된다. 또한 비교 문자에 포함되지 않는 노드를 삭제하고 삭제된 노드의 연결은 상위 노드로 연결을 변경한다. 그림 5(a)에서 노드 B, D, E가 삭제 대상이라고 했을 때 그림 5(b)와 같이 삭제된 노드의 연결을 상위 노드와 연결한다.

3.3 Levenshtein 거리(LD)를 이용한 유사도 측정

LD는 두 개의 비교대상 문자의 유사도를 측정하는 알고리즘으로 문자열의 교정에 사용되는 최소 거리를 계산하게 된다. 하지만 많은 양의 잡음 및 분할이 발생

하면 처리 결과의 신뢰도가 떨어지게 된다. 본 논문에서는 이러한 단점을 보완하고자 LD의 연산을 분석하여 인식 결과에서 잡음 영역과 문자영역을 분리하고 문자 영역에 대하여 문자열거리를 계산하였고 비교 대상 문자와의 유사도를 측정하게 된다. LD algorithm은 입력 문자와 비교문자의 변환행렬을 계산하고 최소 변환 거리를 탐색하며 삽입, 삭제, 대치, 변화 없음으로 구성된 연산을 수행한다.

잡음 영역과 문자 영역을 구분하기 위하여 우선적으로 LD를 수행하게 된다. 입력 문자열조합 X=(E K T O N E H A V E T A J), 비교 대상 문자열 Y=(S T O N E H A V E N)일 때 X와 Y의 LD를 구하기 위한 거리 계산 방법은 그림 6과 같이 표현할 수 있다. 여기서 교정 연산을 분석하여 잡음 영역과 문자 영역을 구분한다. 그림 6의 교정 연산을 분석해보면 양 끝에 삭제 연산이 수행되는 것을 볼 수 있으며 이러한 영역을 잡음으로 간주하고 실제 교정 대상 문자열을 획득하게 된다. 잡음 영역을 제거한 문자 영역 X'=(K T O N E H A V E T)가 되며 LD는 2이고 최장공통부분 스트링의 길이는 5, 공통부분 문자의 수는 7 이다. 여기서 최장공통부분 스트링과 공통부분 문자의 수, LD 교정값을 이용하여 유사도를 판별하게 된다.

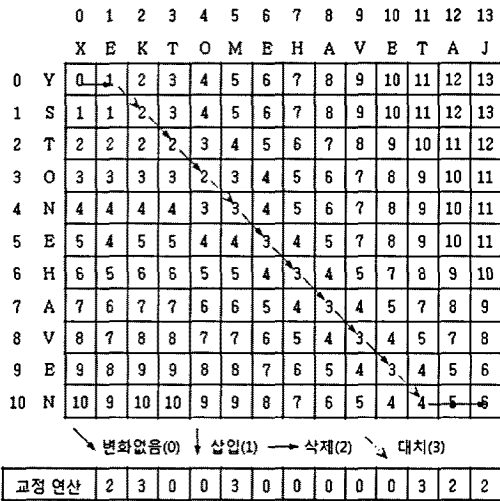


그림 6 Levenshtein 거리

유사도는 식 (1)과 같이 공통부분 문자의 수와 비교 대상 문자의 길이의 비로 표현할 수 있다. 여기서 유사도가 동일한 비교대상 문자에 대해서는 최장공통부분 스트링의 길이를 기준으로 정렬을 수행하여 후처리 결과로 추출하게 된다.

$$\text{유사도} = \frac{\text{공통부분 문자 수}}{\text{비교 대상 문자의 길이}} \times 100 \quad (1)$$

4. 실험 및 결과

실험은 모바일 폰에서 수행하였으며 필기체가 존재하는 18개의 와인 라벨을 대상으로 3명이 각기 라벨 당 3회씩 반복 수행하였다. 실험에 사용된 총 영상 수는 필기체 162장이며 문자수는 필기체 1,647자이다. 실험에 사용된 모바일 폰은 삼성 SPH-M490(T옴니아)를 사용하였다.

실험에 대한 평가 척도로는 후보 문자조합의 축소율, 오인식된 경우를 올바르게 교정한 정도(식 (2)), 후처리 전과 후의 인식률(식 (3))로 하였다.

$$\text{교정률} = \frac{\text{교정문자수} - \text{교정중 오인식수}}{\text{오인식 문자수}} \times 100 \quad (2)$$

$$\text{인식률} = \frac{\text{인식성공문자수}}{\text{총문자수}} \times 100 \quad (3)$$

표 2는 인식결과 필터링을 이용한 문자조합의 축소율로 98.5% 문자조합을 필터링하였고 결과적으로 후처리 과정에서 발생하는 시간 증가를 최소화하였다.

표 2 문자조합의 축소율

	필터링 전	필터링 후	감소율
조합문자 수	18,827	275	98.5%

표 3은 교정률, 표 4는 인식률을 측정한 결과이다. 여기서 라벨단위 인식률은 라벨명을 정확히 인식하였을 경우이며 후처리 전 인식률은 잡음이 포함되어 있는 영상이 대부분이므로 잡음이 포함되어 있다라도 인식 결과에 정답이 모두 존재할 경우 인식 성공으로 하였다.

표 3 문자 단위와 라벨 단위 교정률

	문자 단위	라벨 단위
교정률	73.7%	73.1%

표 4 인식 성능 분석

	라벨 단위		문자 단위	
	정인식 수	인식률	정인식 수	인식률
후처리 전	75	46.3%	865	52.5%
후처리 후	139	85.8%	1,441	87.5%
증감	64	39.5%	3,568	35%

그림 7은 과분할 방식을 이용한 필기체 인식 방법의 결과에 대한 예이다.

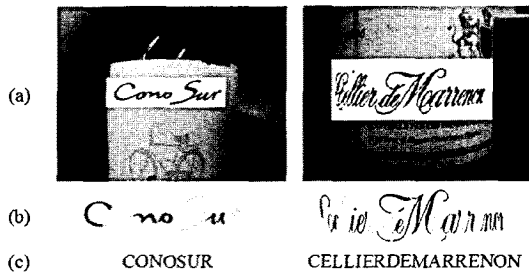


그림 7 필기체 인식의 예 (a) 입력영상, (b) 분할 영상, (c) 인식 결과

5. 결론

본 논문에서는 모바일 폰에서 오프라인 필기체 과분할 인식의 후처리 방법에 관하여 논하였다. 과분할 인식 방법의 인식 결과를 바탕으로 그래프를 구축하고 비교 대상 문자열과 bigram을 비교하여 불필요한 노드의 제거를 수행하였다. 또한 Levenshtein 거리의 교정 거리를 계산하고 교정 연산을 분석하여 잡음영역과 문자영역을 구분하여 문자 영역에서만 거리를 계산함으로써 잡음의 영향을 제거하여 오인식을 개선하기 위한 후처리를 제안하였다. 평가 척도로는 후보 문자조합의 축소율, 오인식된 경우를 올바르게 교정한 정도, 후처리 전과 후의 인식률을 측정하였고 문자 단위 인식률이 39.5%, 라벨 단위 인식률 35.0%의 인식 성능 향상되어 제안한 방법이 효과적임을 보였다. 하지만 문자열이 유사한 경우, 정답의 수가 작은 경우에는 정확한 교정을 수행할 수 없었으며 이러한 문제점도 함께 대처할 수 있는 후처리 방법에 대한 연구도 수행되어야 할 것이다.

참고 문헌

- [1] I. Haritaoglu, "Scene text extraction and translation for handheld devices," *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, vol.2, pp.408-413, 2001.
- [2] C. Thillou, B. Gosselin, "Natural scene text understanding," *Vision Systems, Segmentation and Pattern Recognition*, Ch.16, pp.307-333, 2007.
- [3] J. Zhang, A. Hanneman, A. Hanneman, J. Yang, A. Waibel, "A robust approach for recognition of text embedded in natural scenes," *International Conference on Pattern Recognition*, vol.3, pp.204-207, 2002.
- [4] C. Li, X. Ding, and Y. Wu, "Automatic text location in natural scene images," *Proc. International Conference on Document Analysis and Recognition*, pp.1069-1073, 2001.
- [5] I. S. Oh, C. Y.Suen, C. Y.Suen, "Instance Features for Optical Character Recognition," *Journal of KIISE*

: *Software and Applications*, vol.25, no.7, pp.1028-1043, 1998. (in Korean)

- [6] M. Shridhar, A. Badreldin, "A high-accuracy syntactic recognition algorithm for handwritten numerals," *IEEE Transactions on Systems, Man, and Cybernetics*, vol.SMC-15, no.1, pp.152-158. 1985.
- [7] T. Pavlidis, F. Ali, "Computer recognition of handwritten numerals by polygonal approximations," *IEEE Transactions on Systems, Man, and Cybernetics*, vol.SMC-5, pp.610-614, 1975.
- [8] K. Abend, "Compound decision procedure for unknown distribution and for dependent state of nature," *Pattern Recognition*, pp.207-249, 1968.
- [9] P. Hall and G. Dowling, "Approximate string matching," *Computing Surveys*, pp.381-402, 1980.