

# Client/Server 구조를 이용한 PDA기반의 문자 추출 시스템

박안진<sup>o</sup> 정기철  
숭실대학교, 정보과학대학, 미디어학과  
kcjung@ssu.ac.kr

## PDA-based Text Localization System Using Client/Server Architecture

Anjin Park<sup>o</sup> Keechul Jung  
School of Media, College of Information Science, Soonsil University.

### 요 약

PDA에서 사용하는 대부분의 CPU는 실수 연산 구성요소(float computation component)가 없는 정수(integer) CPU를 사용한다. 인공 신경망(neural network)과 같은 실수 연산이 많은 알고리즘은 PDA에서 많은 수행시간을 가진다. 본 논문에서는 이런 단점을 해결하기 위해 무선 랜(LAN)으로 연결된 Client(PDA)/Server(PC) 구조를 이용한 효과적인 문자 추출 시스템을 제안한다. Client(PDA)는 대략적인 문자 추출 결과를 JPEG으로 압축하여 전송속도를 최소화한다. Server(PC)는 Client(PDA)의 결과를 바탕으로 정밀한 문자 영역 추출을 위해, 텍스처 분류 방법과 연결 성분 분석 방법을 이용한다. 실험에서 제안한 방법은 속도뿐만 아니라 문자 추출에서도 효과적이었다.

### 1. 서론

언제, 어디서, 누구나 대용량 통신망을 사용할 수 있는 유비쿼터스(ubiquitous) 시대가 다가오면서 카메라가 장착된 무선 통신이 가능한 PDA와 같은 휴대용 장치가 가까운 미래에는 일상의 한 부분이 될 것이다.

문자기반 영상 인덱싱에서의 문자 추출은 자동 문자 인식(OCR)의 전 단계로써 중요하다. 또한 영상내의 문자 추출에 관한 연구가 멀티미디어 시스템, 전자 도서관, 비디오 인덱싱, 문서 구조 분석, 우편 영상내의 주소 영역 추출, 자동차 번호판 추출 등 다양한 관련 분야에서 진행 되고 있으며, 휴대용 장치에서도 이와 관련된 많은 문자 추출 및 기타 연구가 진행되고 있다 [1-6].

PDA에서 문자 추출 및 인식에는 일반적인 PC와 달리 크게 두 가지 문제점을 가지고 있다 [4].

첫째, 제한된 연산 자원이다. 대부분의 PDA는 실수 연산 구성 요소(float computation component)가 없는 정수(integer) CPU를 사용하며, 실수 연산을 수행하기 위해서는 실수 대항

라이브러리(float emulation library)를 사용하므로 오랜 수행시간이 소요된다. 예를 들어, 본 논문에서 사용한 MLPs(multi-layer perceptrons)기반의 분류 방법을 이용하여 240×320 영상에 대해 PDA에서 문자 추출을 수행하면 4분 정도의 수행시간이 소요되며, 이는 본 실험에서 사용한 PC에 비해 120배 느리다 [1].

둘째, 제한된 저장공간(memory)이다. 예를 들어 본 논문에서 사용한 Pocket PC 2002기반의 PDA는 16~64MB 저장공간을 가지며, Palm OS기반의 PDA는 보통 8~16MB 저장공간을 가진다. 이는 문자 인식을 올리기에 너무 작은 공간이다.

이런 문제점을 해결하기 위해 Zhang [4] 등은 부동 연산을 일반화된 정수연산으로 바꾸어 문자 추출 하는 방법을 사용하였다. 그러나, 이 방법은 수행속도는 빨라졌으나 정확도 손실을 가져왔다. Haritaoglu [5]는 이런 문제점을 해결하기 위해 Client/Server 구조를 이용하여, 부동 연산이 많이 수행되는 정밀한 문자 영역 추출은 Server(PC)에서 수행 하는 방법을 사용하였다. 그러나, PDA에서 사용자가 수작업으로 대략적인 문자 영역을 지정하여야 하며, 무선 랜(LAN)에서의 전송 속도 역시

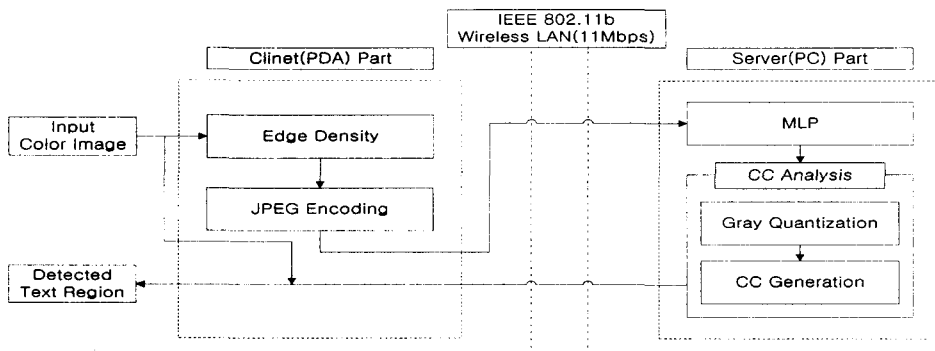


그림 1. 제안된 문자 추출 시스템 구성도.

문제가 되었다.

그림 1은 본 논문에서 제안하는 시스템을 나타내고 있다. 우리는 제한된 연산 자원, 제한된 저장 공간과 같은 PDA의 문제점을 해결하기 위해 Client/Server구조를 이용하며, Client/Server 사이의 무선 링크를 통한 전송 속도 문제를 해결하기 위해, PDA에서 검출된 대략적인 문자 영역을 JPEG 알고리즘을 이용하여 영상을 압축, 전송한다. Server(PC)에서는 Client(PDA)의 결과를 바탕으로 정밀한 문자 영역 추출을 수행하며, 저 해상도에서도 높은 검출률(recall rate)을 가지는 MLPs(multi-layer perceptrons)기반 방법과 정확률(precision rate)을 높이기 위한 연결 성분(connected component: CC) 분석기반의 문자 추출 방법을 동시에 사용한다.

본 논문의 구성은 다음과 같다. 제 2장에서 Client(PDA)에서 수행하는 방법을 기술하고, 제 3장에서는 Server(PC)에서 수행하는 문자 추출 방법을 기술하며, 제 4, 5장에서 실험결과 및 향후 연구 방향을 기술한다.

## 2. Client(PDA)

Client(PDA)는 에지 밀도(edge density)를 이용하여 대략적인 문자 영역을 추출 하고, 결과 영상을 압축하여 전송 속도를 최소화한다.

우리는 대략적인 문자 영역을 추출하기 위해 에지 밀도를 이용하였으며, 비-문자 영역은 0으로 채운다.

에지 밀도는  $N \times N$  Window 안에서의 에지 포인트(Edge Point) 개수로 말하며, 에지 밀도가 임계 값(threshold) 이상이면 대략적인 문자 영역으로 생각한다.

본 논문에서는 영상 압축 방법으로 JPEG을 이용한다. JPEG 알고리즘은 효율적인 압축률을 보이며, 1/20까지 압축률을 높여도 원래의 영상과 비교해서 큰 차이가 없으며, 연속적인 색조 영상의 처리에 효과적이다[7].

라플라시안(laplacian)을 이용하여 에지 포인트를 구하며, Window Size로  $20 \times 20$ 을 사용하였고, 임계 값으로 30을 사용하였다.



그림 2. 에지 밀도를 이용한 결과 이미지.

Server(PC)에서는 그레이 영상(gray image)만으로 문자 영역 추출을 수행하므로, 에지 밀도의 결과 영상을 그레이 영상으로 바꾸어서 압축, 전송한다.

그림 2는 (a) PDA에서의 입력 영상과 (b) 에지 밀도를 이용한 결과 영상을 나타낸다.

## 3. Server(PC)

Server(PC)는 Client(PDA)의 결과를 바탕으로 정밀한 문자 영역 추출을 수행하며, MLPs기반의 텍스트 분류 방법과 연결 성

분 분석기반의 문자 추출 방법을 이용한다.

PDA에서 입력 받은 영상은 저해상도의 영상이며 다양한 배경이 포함될 수 있다. MLPs 기반의 텍스트 분류 방법을 이용하면 이런 영상에서도 높은 검출률을 보였으며, 정확률을 높이기 위해 연결 성분 기반의 문자 추출 방법을 같이 사용하였다.

## 3.1 MLPs 기반의 문자 추출 방법

문자 영역과 비-문자 영역을 구분하는 텍스트 분석기를 구성하는 어려움을 극복하기 위해 학습을 통한 텍스트 분류기 생성에 관한 연구가 진행되고 있으며[3, 6], 본 논문에서는 MLP를 사용하여 다양한 크기나 모양의 문자와 배경에 적응할 수 있는 텍스트 분류기를 구성한다. MLP는 2개의 은닉층, 1개의 출력층으로 구성되며 인접층의 노드들은 모두 연결되어 있고, 입력층은 그림 3과 같이, 입력 영상에서  $M \times M$  크기의 입력창 내의 검은색으로 표시된 화소들의 인텐시티 값을 사용한다. 그림 3은 신경망의 입력층이 생략되어 있고, 대신 입력 화소의 모양으로 이를 대신한다. 이러한 입력 구조는 텍스트 분석 분야에서 성능과 속도 향상에 좋은 것으로 알려져 있다[6]. MLPs를 이용하여 입력 영상을 스캔함으로써 생성된 결과 영상은 입력 영상의 각 화소를 문자와 비-문자 클래스로 구분한다.

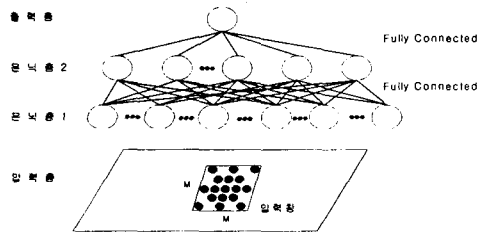


그림 3. 신경망의 개략적인 구조.

## 3.2 연결 성분 기반의 문자 추출 방법

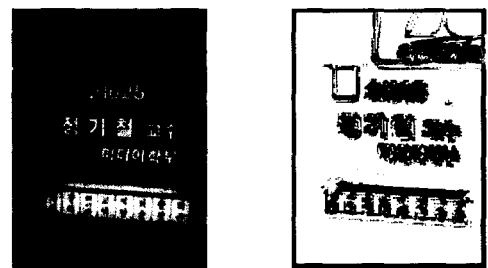


그림 4. MLP를 이용한 문자 추출에 (a) 문자 영역을 표시한 입력 이미지, (b) MLP 수행 결과.

MLPs기반의 문자 추출 방법은 가능한 많은 문자 영역 추출을 원하기 때문에 잘못된 추출(false alarm)을 포함 할 수 있으며, 정확률이 떨어질 수 있다. 그림 4는 MLPs 기반 문자 추출을 수행한 결과를 나타내며, 고-대비(high-contrast) 또는 고-주파(high-frequency) 영역과 문자와 비슷한 텍스트 특성을 가진 영역에서 많은 잘못된 추출이 있음을 보여준다. 우리는 MLPs 기반 방법의 이런 문제를 해결하기 위해 연결 성분 기반의 문자 추출 방법을 사용하였다.

본 논문에서는 연결 성분을 만들기 위해 MLPs의 결과영상을 3비트로 양자화를 하였으며, 다음 두 단계의 필터링을 적용하여 정확도를 높였다.

단계 1: 크기, 가로, 세로의 길이와 같은 연결 성분의 특징을 이용: 그림 8에서 보는 바와 같이 저 해상도 영상에서는 글자마다 연결 성분이 형성되지 않는다. 본 논문에서는 MAX\_크기, MAX\_가로, MAX\_세로와 같이 미리 정한 연결 성분의 특징 값보다 큰 연결 성분에 대해서만 필터링을 실행한다.

단계 2: 기하학적 배열: 같은 색깔을 가진 근접한 연결 성분의 수를 확인한다. 문자는 3개 이상의 연속된 연결 성분을 가져야 한다.

그림 5는 MLPs의 수행 결과에서 연결 성분의 두 단계 필터링을 적용한 결과를 보여준다.

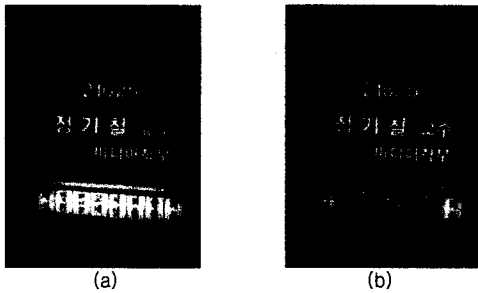


그림 5. 연결 성분 필터링에 대한 결과

(a) MLPs 기반의 수행 결과, (b) 연결 성분 기반의 수행 결과.

#### 4. 실험 결과

본 논문에서 사용한 PDA는 Pocket PC 2002가 장착된 HP iPAQ h5450 모델을 사용하였으며, 카메라는 PDA용 everCam LMC2201B를 사용하였다. Server(PC)는 Pentium IV 2.66GHz 하드웨어 상에서 구현하였다.

실험에 240×320크기의 영상을 사용하였으며, 그림 6, 7은 Server(PC)에서의 단계별 결과를 보여주며, 표 1은 단계별 수행 시간을 나타낸다.

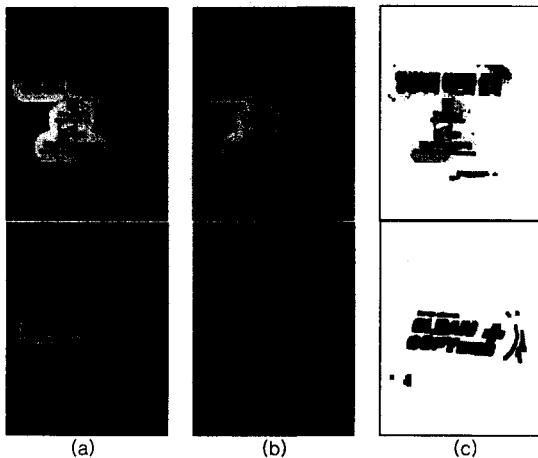


그림 6. 서버에서의 단계별 결과

(a) Client(PDA)에서 입력 받은 영상(에지 밀도 결과)  
(b) MLPs 기반의 수행 결과 (c) 연결 성분 기반의 수행 결과.

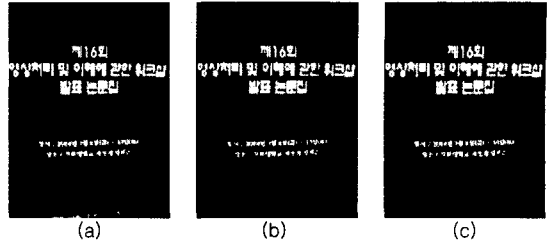


그림 7. 서버에서의 단계별 결과.

표 1. 단계별 수행 시간.

단계	수행 시간(ms)
에지 밀도	180
전송 시간	30
MLPs 기반 방법	2300
연결 성분 기반 방법	250
총 수행 시간	2760

#### 5. 결론

본 논문에서는 PDA에 장착된 카메라로부터 입력 받은 영상을 이용하여 문자 추출 시스템을 기술하였으며, 속도뿐만 아니라 문자 추출에서도 효과적이었다.

우리는 이 시스템을 개발하면서 하나의 문제점에 직면하게 되었다. PDA에서 입력 받은 영상은 저해상도의 영상으로, 그림 8에서 보는 바와 같이 글자마다 연결 성분이 형성되지 않았다. 이 문제는 글자마다 연결 성분이 구성될 수 있는 전 처리과정 에 관한 연구 또는 고 해상도 카메라를 이용하여 해결할 수 있다.



그림 8. 8bit 양자화 결과.

#### 6. 참고 문헌

- [1] 박만진, 정기철, " PDA를 이용한 문자 추출 시스템," 16th Image Processing Image Understanding, pp. 277-280, 2004.
- [2] Keechul Jung, Kwang In Kim, and Anil K. Jain, " Text Information Extraction in Image and Video: A Survey," International Journal of Pattern Recognition, To be published.
- [3] Keechul Jung, " Neural Network-based Text Location in Color Images," Pattern Recognition Letters Vol. 22, No. 14, pp. 1503-1515, 2001.
- [4] Jing Zhang, Xilin Chen, Jie Yang, Alex Waibel, " A PDA-based Sign Translator," Fourth IEEE International Conference on Pattern Recognition, pp. 216-219, Oct. 2002.
- [5] Ismail Haritaoglu, " Scene Text Extraction and Translation for Handheld Devices," IEEE Computer Society Conference on, Vol. 2, 8-14, Dec. 2001.
- [6] Anil K. Jain and Kalle Karu, " Learning Texture Discrimination Masks," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 18, No. 2, pp. 195-205, 1996.
- [7] Randy Crane, " A Simplified Approach to Image Processing," Prentice Hall PTR, 1997.