

주파수 특성과 역전파 신경망 알고리즘을 이용한 문자 영역 분할 방법

전병태*, 송치양**

A method for Character Segmentation using Frequency Characteristics and Back Propagation Neural Network

Byung Tae Chun*, Chee Yang Song**

요약

문자 영역 추출을 위해서 FFT와 신경망을 이용한 방법을 본 논문에서 제안하고자 한다. 일반적으로 문자 영역은 고주파 영역에서 발견되므로 FFT를 이용하여 이 특징을 추출할 수 있다. 문자(고 주파) 영역과 비 문자(저 주파) 영역을 신경망에 학습을 시킨다. 신경망에 고주파 영역을 입력으로써 후보 영역을 추출한다. 그리고 최종 문자 영역은 후보 영역 검증을 통하여 추출된다. 실험 결과 후보 영역 추출은 학습된 경우 100% 추출율을 보여주고 있으며, 검증을 통한 후보 영역 추출율은 95%임을 알 수 있었다. 제안된 알고리즘의 장점은 알고리즘의 단순성과 실시간 처리에 있다.

Abstract

The proposed method uses FFT(Fast Fourier Transform) and neural networks in order to extract texts in real time. In general, text areas are found in the higher frequency domain, thus, can be characterized using FFT. The neural network are learned by character region(high frequency) and non character region(low frequency). The candidate text areas can be thus found by applying the higher frequency characteristics to neural network. Therefore, the final text area is extracted by verifying the candidate areas. Experimental results show a perfect candidate extraction rate and about 95% text extraction rate. The strength of the proposed algorithm is its simplicity, real-time processing by not processing the entire image.

▶ Keyword : 비디오 영상(video image), 문자 영역 추출(Text extraction)

-
- 제1저자 : 전병태
 - 접수일 : 2006.07.24, 심사일 : 2006.08.13, 심사완료일 : 2006.09.20
 - * 국립 한경대학교 웹정보공학과 교수
 - ** 국립 상주대학교 소프트웨어공학과 교수

I. 서론

최근 컴퓨터 하드웨어의 급속한 발전으로 처리 속도 향상과 메모리의 대용량화가 가능하게 됨으로서 과거 컴퓨터 환경의 제한으로 개발되지 못했던 많은 기술들의 개발이 가능하게 되었다. 컴퓨터 환경의 향상과 멀티미디어 기술의 급속한 발전으로 대용량 멀티미디어 저장/서버 기술이 발전하게 되었으며 그 기술 중 하나가 영상/동영상 데이터 베이스 기술이다.

기존의 영상/동영상 데이터 베이스 시스템은 단순한 키워드에 의하여 검색함으로써 다양한 사용자의 검색 요구에 부응할 수 없는 단점이 있었다. 이러한 문제점을 해결하기 위하여 최근에 내용기반 동영상 검색 시스템 개발에 많은 관심을 갖게 되었다. 내용 기반 동영상 검색 시스템은 추출된 영상의 특징 정보(컬러, 모양, 자막, 텍스트, 건본 영상, 등)를 메타 데이터화하여 데이터 베이스에 저장하고, 이 특징에 대하여 질의함으로써 질의 특징과 가장 유사한 영상을 추출하여 주는 시스템이다 [1][2][3]. 이들 특징 정보 중 영상의 의미를 가장 함축적으로 표현해놓은 의미 정보는 비디오에 존재하는 자막 정보일 것이다. TV 프로그램에서 뉴스, 여행안내, 취미, 교육 훈련에 관한 문자 정보는 대표적인 영상의 의미 정보가 될 수 있다. 따라서 이러한 문자 정보를 자동적으로 추출하고 인식하여 의미 화할 수 있는 연구가 필요하다.

문자 영역 추출에 관한 기존의 연구[4][5][6]는 공장 자동화, 사무 자동화, 산업 자동화에서 필요로 하는 문자 추출 및 인식에 관한 연구가 주로 이루어 졌다. 이들 시스템은 어느 정도 좋은 연구 결과가 이루어져 여러 분야에서 일부 사용되고 있다. 그러나 이 연구들은 문서 영상과 같은 이진 영상을 처리하기 위한 방법으로 명암도(gray value)나 컬러 영상과 같은 영상을 처리하기에는 부적합하다고 볼 수 있다.

기존의 문자 영역 추출 방법은 자막을 텍스트로 간주하고 가보 필터링[7]과 같은 텍스트적 분석 방법[8]을 이용하거나 공간적 변화 분석 방법[9]을 이용하여 자막 영역을 추출하였다. 위와 같은 텍스트 분석 방법은 문자의 크기 및 서체에 민감하고 문자 영역을 정확히 추출할 수 없을 뿐 아니라 처리 시간이 많이 소요된다는 단점이 있다.

전체 영상에 대하여 영역 분할 및 합병(region split

and merge) 방법[10]을 이용하여 문자 영역을 추출하거나, 프레임간의 차를 이용한 문자 영역 추출 방법[11]들은 문자에 대한 사전 지식이 필요하며 전체 영상을 처리함으로써 처리 시간이 많이 소요된다는 단점이 있다. 컬러의 축소(color reduction) 방법[12][13]은 컬러 히스토그램, 컬러 양자화(quantization), 컬러 비트 축소(color bit reduction) 기법들을 이용하여 문자 영역을 추출한다. 이러한 방법들은 추출될 컬러의 사전 지식(문자의 색상, 영역 비율 등)이 필요로 한다는 문제점이 있다. 그 이외에도 비디오에 존재하는 3차원 문자를 추출하기 위한 방법[14]도 제안되기도 하였다.

II. 문자 영역의 주파수 특성

문자 영역의 명암도 특성을 보면 그림1과 같이 문자 부분과 비 문자 부분의 명암도가 다름을 볼 수 있다. 즉 문자 부분은 주기적인 주파수 형태가 발생하고 비 문자 영역은 불규칙한 주파수 형태를 보여주고 있다[15]. 문자 부분의 규칙성을 주파수 영역으로 변환하면 일정한 형태의 주파수 특성을 갖는 것을 볼 수 있다. 주파수 변환을 위해서 본 논문에서는 FFT 변환 방법을 이용하였다. 문자 영역과 비 문자 영역에 대한 FFT 수행 결과를 보면 그림 2, 3과 같다.

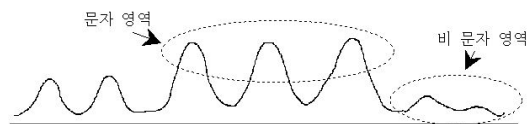


그림 1. 문자 및 비 문자 영역의 명암도 특성
Fig.1 contrast characteristic of region for text and non-text

그림 2 (a) 상단 1*64의 샘플링에 대한 주파수 특성을 보면 그림(b)와 같다. (a) 부분은 비 문자 부분으로 저주파가 형성됨을 볼 수 있다. 그림 2 (c) 중단 1*64의 샘플링에 대한 주파수 특성을 보면 그림(d)와 같다. (c) 부분은 문자 부분으로 고주파가 형성됨을 볼 수 있다. 그림 2(e) 상단 1*64의 샘플링에 대한 주파수 특성을 보면 그림(f)와 같다. (e) 부분은 비 문자 부분으로 저주파가 형성됨을 볼 수 있다.

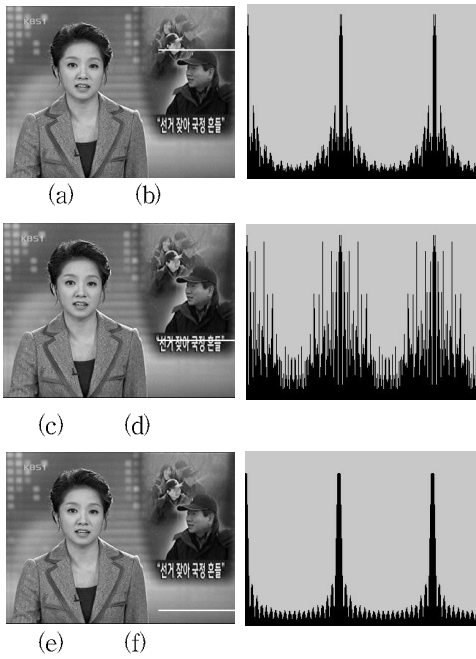


그림 2. 문자 및 비 문자 영역의 주파수 특성
Fig2. frequency characteristic of region for text and non-text

그림 3 (a) 상단 1*64의 샘플링에 대한 주파수 특성을 보면 그림(b)와 같다. (a) 부분은 비 문자 부분으로 저주파가 형성됨을 볼 수 있다. 그림 3 (c) 중단 1*64의 샘플링에 대한 주파수 특성을 보면 그림(d)와 같다. (c) 부분은 문자 부분으로 고주파가 형성됨을 볼 수 있다. 그림 3 (e) 하단 1*64의 샘플링에 대한 주파수 특성을 보면 그림(f)와 같다. (e) 부분은 비 문자 부분으로 저주파가 형성됨을 볼 수 있다.

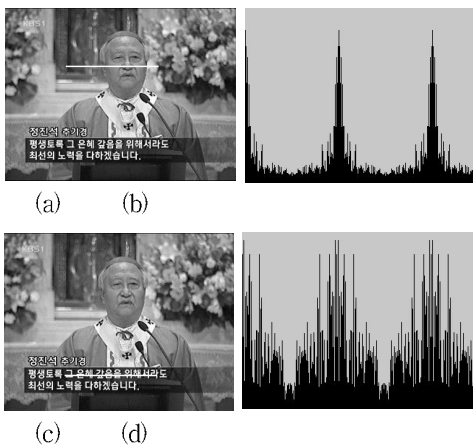
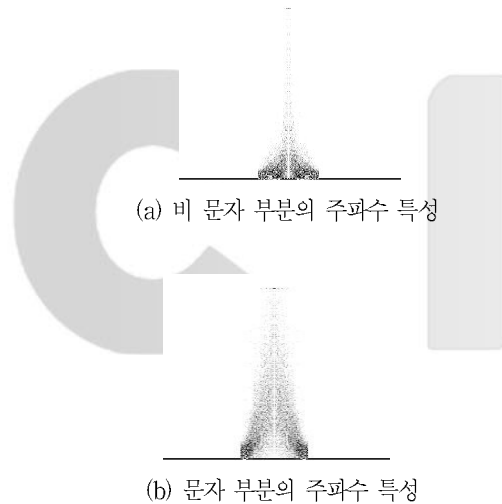


그림 4. 주파수 특성
Fig4. frequency characteristic



본 논문에서는 이러한 주파수 특징을 신경망 학습을 통하여 문자 영역을 추출하고자 한다. 신경망 학습을 위하여 수집된 문자/비 문자 영역의 주파수 특성을 종합하여 보면, 비 문자 영역에서는 저주파가 형성되고 단조 감소함을 볼 수 있으며, 문자 영역에서는 고주파가 형성되고 주파수의 크기도 큼을 볼 수 있다.

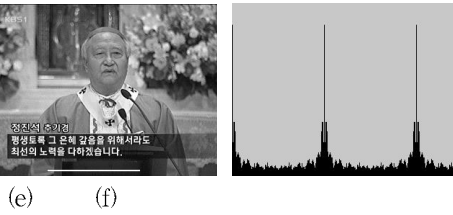


그림3. 문자 및 비 문자 영역의 주파수 특성
Fig.3 frequency characteristic of region for text and non-text

문자 및 비 문자 영역에 대한 FFT 특성을 종합적으로 살펴보면 그림 4(a),(b)와 같다. 그림4 (b)에서 보여지는 비 문자 부분의 주파수 특성은 저주파 부분 많이 형성되고 단조 감소함을 볼 수 있으며 주파수의 값이 작은 것을 볼 수 있다. 반면 그림4 (a)에서 보여지는 문자 부분의 주파수 특성은 고주파 부분이 크게 형성되고 주파수 값이 큰 것을 볼 수 있다.

III. FFT와 역전파 신경망을 이용한 문자 후보 영역 추출

문자 후보 영역 추출을 위하여 얼마만큼의 주파수 특성을 구할 것인가 중요한 문제가 될 수 있다. 다시말해 얼마만큼의 화소를 입력하여 FFT 주파수 특성을 산출하는 문제이다. FFT에 입력하는 화소가 많을(길) 경우 문자 영역과 비 문자 영역이 혼합되어 문자/비 문자 영역 구분에 변별력이 없는 문제점이 있고, FFT에 입력하는 화소가 적을 경우 문자 영역과 특성이 비슷한 비 문자 영역이 문자 영역으로 나올 가능성이 많은 문제점이 있다.

본 논문에서는 FFT 특성을 구하기 위하여 $1 \times 64(2n, n=6)$ 의 명암값을 사용한다. 그 이유는 $128(2n, n=7)$ 의 경우 너무커서 비 문자 영역과 문자 영역이 동시에 존재할 가능성이 많기 때문이다. FFT의 명암 값 $64(2n, n=6)$ 화소는 문자 영역이 작은 규모의 문자 영역도 포함할 수 있기 때문에 64 화소가 적정하다고 판단한다. FFT의 입력으로 64가 주어진다면 나이퀴스트 주파수 이론에 의하여 실제 주파수의 출력은 32가 된다. 주어진 신경망의 입력 개수는 32개이며 은닉층은 40개의 뉴런을 사용했다. 출력 층은 1개로 문자 영역 비 문자 영역으로 구분한다. 여러 가지 비디오에서 학습을 위한 샘플링을 수행한다. 학습을 위한 데이터 추출 방법은 그림 5와 같다.



그림 5. 주파수 입력을 위한 샘플링 구간
Fig.5 sampling range for input data

위와 같은 샘플링 구간의 명암도 값을 신경망에 입력하여 문자 후보 영역 구간인지를 판별한다. 판별된 후보 영역을 상호 연결(영역 형성)을 통하여 그림 6과 같이 문자의 후보 영역을 추출한다.



그림 6. 추출된 후보 영역
Fig.6. extracting candidate region of text

IV. 추출된 문자의 후보 영역 검증

후처리 방법은 문자의 후보 영역이 문자인지 아닌지 결정하기 위하여 사용한다. 후보 영역 검증하기에 앞서 문자 영역으로 보기 힘든 영역을 제거하는 잡음 제거 단계를 거친다. 문자 영역에 대하여 리벨링 연산을 수행하고 구해진 각 영역이 문자 영역으로 보기에 너무 크거나 작은 영역을 제거한다. 이러한 전처리 단계를 수행함으로써 잡음 영역이 검증단계에 참여하는 것을 방지 할 수 있다. 그림 7는 검증 과정을 보여주고 있다. 문자 영역을 추출하기 위하여 2단계 검증 과정을 거친다. 첫 번째 단계는 후보 문자 라인의 여러 특징을 비교하여 비 문자 영역을 제거한다. 즉, 문자 영역의 크기, 문자 영역의 가로 세로가 너무 작거나 문자로 보기에 너무 큰 영역은 제거한다. 다음으로 문자 영역의 가로 세로 비를 계산하여 이 비가 너무 작거나 너무 크면 이 영역을 제거한다. 그리고 문자 영역의 흰 화소(문자) 대 검정 화소(배경)의 비가 너무 적으면 이 영역 역시 제거한다. 두 번째 단계에서는 추출된 문자 영역의 대단위 영역화에서 문자 영역 안에 있는 문자의 수가 3자 이하 인 것은 문자 영역으로 간주하지 않는다.

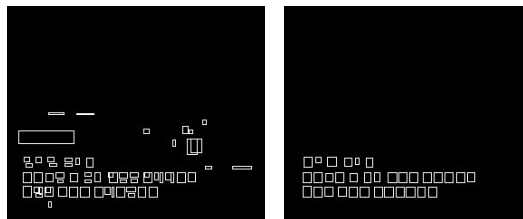


그림 7. 후보 문자 영역의 검증
Fig.7. verifying candidate region of text

그림 8은 검증 후 추출된 문자 영역을 보여주고 있다.



그림 8. 검증 후 추출된 문자 영역
Fig.8. extracted text region after verifying

V. 실험 결과 및 고찰

사용된 프로그램은 C++6.0이며 컴퓨터는 펜티엄 PC 3.0GHz CPU를 사용했다. 실험을 위하여 MPEG1 영상을 사용했으며 TV 영상을 영상 캡처 보드를 사용하여 MPEG 영상을 획득했다. 비디오 영상은 주로 뉴스를 사용했고 초당 3 프레임 간격으로 실험 영상을 사용했다.

실험 영상은 영화, 뉴스, 드라마 등을 사용했다. 실험 결과 표 1과 같이 뉴스 4472 프레임에 대하여 후보 영역 추출율은 99%, 문자 영역 추출율은 98%임을 알 수 있다. 영화의 경우 4279 프레임에 대하여 문자 영역 추출율은 99%, 문자 영역 추출율은 97%임을 알 수 있다. 스포츠의 경우 3634 프레임에 대하여 후보 영역 추출율은 98%, 문자 영역 추출율은 97%임을 알 수 있었다. 실험 결과 FFT를 이용한 문자 영역 추출 방법은 매우 효과적임을 볼 수 있었다.

표 1. 문자 영역 추출 결과
Table1. result for extracting text area

	프레임 수	문자 후보 영역 추출율(%)	문자 영역 추출율(%)
뉴스	4472	99	98
영화	4279	99	97
스포츠	3634	98	97

기존의 휴리스틱을 이용한 문자 인식 방법은 여러 가지 환경에 따라 파라미터 조정이 필요하나 신경망 학습에 방법은 추출될 문자의 특성을 반영하여 학습하기 때문에 타 방법 보다 문자 영역 추출율이 비교적 높다고 볼 수 있다.

VI. 결 론

본 논문에서는 FFT와 신경망을 이용한 비디오 영상에서 문자 영역을 추출하는 방법을 소개하였다. 많은 실험 결과에서 그 성능을 입증할 수 있었다. 그러나 향후 다음과 같은 부분에 대하여 연구를 진행하고자 한다. 제안한 방법이 배경 영역과 문자 컬러 영역이 명암도 차가 크지 않을 경우 효과적이지 못 할 수 있다. 이러한 명암도 차가 적은 부분에서 효과적인 문자 후보 영역 검증 방법에 대하여 개선하고자 한다.

참고문헌

- [1] HongJiang Zhang, Shuang Yeo Tan, Stephen W. Smoliar and Gong Yihong, "Automatic parsing and indexing of news video," Multimedia System, Vol.2, pp.256-266, 1995.
- [2] HongJiang Zhang, C.Y.Low, S.W.Smoliar and J.H.Wu, "Video Parsing, Retrieval and Browsing : An Integrated and Content-based Solution," Proc. ACM Multimedia 95, San Francisco, CA, pp.15-24, Nov. 1995.
- [3] Michael A. Smith, and Takeo Kanade, "Video Skimming for Quick Browsing based on Audio and Image Characterization," Carnegie Mellon Univ., Technical Report CMU-CS-95-186, July 1995.
- [4] Jun Ohya, Akio Shio and Shigeru Akamatsu, "Recognizing Characters in Scene Images," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.16, No.2, pp.214-220, 1994.
- [5] B. Yu and A. Jain, "Robust and fast skew detection algorithm for generic documentation," Pattern Recognition, Vol.29, pp.1599-1629, 1996.
- [6] B. Yu and A.jain and M. Mohiuddin, "Address block location on complex mail piecies," Proc.

of the 4th Int. Conf. on Document Analysis and Recognition, Ulm, 1997.

[7] A. Jain and S. Bhattacharjee, "Text segmentation using Gabor filters for automatic document processing," Machine Vision Application, Vol.5, pp.169-184, 1992.

[8] I. Pitas and C. Kotropoulos, "A textured-based approach to the segmentation of semitic image," Pattern Recognition. Vol.25, pp.929-945, 1992.

[9] Y. Zhong, K. Karu, and A. Jain, "Locating text in complex color images," Pattern Recognition, vol.28, pp.1523-1535, 1995.

[10] Rainer Lienhart and Frank Stuber, "Automatic text recognition in digital videos," Proc. of the SPIE, Image and Video Processing IV, Vol. SPIE2666, pp.180-188, San Jose, 1996.

[11] Shoji Kurakake, Hidetaka Kuwano and kazumi Odaka, "Recognition and visual feature matching of text region in video for conceptual indexing," Proc. of the SPIE, Storage and Retrieval for Image and Video Database V, vol. SPIE3022, pp.368-378, San Jose, 1997.

[12] Yu Zhong, Kalle Karu and Anil K. Jain, "Location text in complex color images," Pattern Recognition, Vol.28, No.10, pp.1523-1535, 1995.

[13] Hae-Kwang Kim, "Efficient automatic text location method and content-based indexing and structuring of video database," Journal of Visual Communication and Image Representation, Vol.7, No.4, pp.336-344, Dec. 1996.

[14] Jun Ohya, Akio Shio and Shigeru Akamatsu, "Recognizing characters in scene images," IEEE Transactions on Pattern Analysis and Machine Intelligenece, Vol.16, No.2, pp.214-220, 1994.

[15] Byung Tae Chun, Younglae Bae and Tai-Yun Kim, "Text Extraction in Videos using Topographical Features of Characters," The 8th IEEE Int. Conf. on Fuzzy System(Fuzz-IEEE'99), Vol.2, pp.1126-1130, Seoul, 1999.

저자 소개



전 병 태

1986년 한남대학교 전산과(학사)
 1989년 숭실대학교 전산과(석사)
 2001년 고려대학교 컴퓨터학과(박사)
 1992년 5월 IR52 장영실 상 수상
 (과기부 장관상)
 1989년~2004년 한국전자통신연구원
 선임연구원
 2004년 2월 ~ 현재 국립 한경대학교
 웹정보공학과 교수
 2003.8~현재 정통부 산하 프로그램
 심의조정위원회 감정전문위원
 2004.7~현재 한국정보기술학회 이사
 2004.5~현재 한국전자통신연구원
 초빙연구원



송 치 양

1985년 한남대학교 전산학과 학사
 1987년 중앙대학교 전산학과 석사
 2003년 고려대학교 컴퓨터학과 박사
 1990년 ~ 2005년 한국통신 중앙연구
 소 책임연구원
 2005년 ~ 현재 상주대학교 소프트웨
 어공학과 전임강사
 관심분야 : UML 모델링 기술, 소프트
 웨어 개발방법, 멀티미디어 영상 처리