

자연 영상에서 문자열 추출

최미화^U 김희승

서울시립대학교 전산통계학과
(mhchoi91, hskim)@venus.uos.ac.kr

Text Location in Scene Images

Mi-Hwa Choi^U Hee-Seung Kim

Dept. of Computer Science and Statistics, University Of Seoul

요 약

본 논문을 자연영상에서 문자열의 위치를 찾아내는데 모폴로지 연산인 WTH(white top-hats)과 BTH(black top-hats)을 사용하였다. 기존의 자연영상에서의 문자열추출은 칼라양자화방법 경우 각 칼라공간에서 문자열추출과정을 반복 적용하거나 모델기반방법의 경우 문자열의 획의 크기나 특징에 따라서 하나의 영상을 여러 개로 분리 적용하여 추가적인 계산비용을 필요로 한다는 점을 개선하고 공간적 변화도를 이용하여 영상을 직접 처리하는 경우 최소 문자열 후보영역을 찾기 위한 프로세스를 다시 적용해야 한다는 점을 개선하였다. 자연영상에 문자열의 위치를 대략적으로 찾아내기 위해 WTH+BTH를 적용하여 그 결과로 문자열의 대략적 위치와 최소문자열후보영역을 동시에 얻을 수 있다. 문자열이 가지는 특성을 적용하여 문자열-비문자열 분류과정을 적용하고 후처리를 통해 완전한 문자열의 위치를 보여준다.

1. 서론

자연영상에는 많은 복잡한 정보들이 존재한다. 그러한 정보 중에서 문자열은 장애인보행시스템, 로봇운행시스템 및 여러 비전 응용 시스템에서 상당히 중요한 정보가 된다.

문자열추출은 지난 20여년 동안에 문서분석분야에서 많은 연구 및 진보가 있어왔다. 주로 단색문서영상에서 페이지 영역화 작업이 주가 되었으나 최근에는 CD 표지, 문헌표지등 칼라 이미지에서 문자열추출을 위한 기술들이 제안되었다. 그러나 문자열이 배경 및 비문자열 부분과 겹치기 쉬운 실제 응용시스템에서는 추출과정이 상당히 어려워진다.

자연영상은 문서영상보다 다양한 폰트, 다양한 크기, 복잡한 배경, 문자열-비문자열 겹침현상, 복잡한 칼라, 영상을 획득한 시점에서 생기는 문자열의 3차원 현상등 추출을 복잡하게 하는 요소들이 많이 존재한다.

2. 관련연구

1980년 초반에 페이지 영역화 작업을 수행하기 위한 많은 기술들이 제안되어 왔다. 그러나 대부분의 기술들은 단색 문서영상을 처리하는 것이었다. 대표적인 영역화 알고리즘은 RLSA(run-length smearing algorithm)과 RXYC(recursive X-Y cuts)이다. RLSA는 1974년 Jahnson[2]에 의해 제안되었고 그래픽으로 부터 문서영역을 구별하기 위한 것으로 1982년 Wong et al.[3]에 의해 확장되었다. 이 방법은 서로 근접한 검은 픽셀을 연결하여 이진 이미지를 연결하는 것이다. 그 다음에

수직과 수평방향으로 적용한 결과를 결합함으로써 블록을 찾아낸다. RXYC은 1984년 Nagy[4]에 의해 문서영상 영역화에 사용하기 위해 제안된 알고리즘으로 수직과 수평으로 영상을 투사하여 깊은 골에 해당하는 부분을 잘라주는 방법이다. 위 두가지 유명한 방법을 이용한 영역화방법들은 심하게 기울어진 각을 가진 문자열에 대해서는 좋은 결과를 얻기 힘들다. 기울어진 짐을 찾기 위해서 허프변환, 가장 가까운 이웃 그룹화방법들이 이용되었다. 문자열 영역을 결코 간주하여 처리하는 기법도 제안되었으나 결기반 영역화 방법은 문자의 크기, 모양에 민감하고 처리시간이 많이 걸린다.

최근에는 복잡한 레이아웃을 가진 칼라 인쇄문이 점점 많아짐에 따라 칼라를 처리하기 위한 기법이 제안되었다. 메디언컷(Median-cut) 칼라 양자화는 칼라의 정보를 줄여서 처리할수 있는 가장 간단하고 유용한 방법이다. 칼라양자화는 각각의 칼라공간에서 영역화를 수행하므로 많은 시간이 들 수 있고 여러개의 칼라가 하나의 대표칼라로 설정함으로써 불필요한 정보가 생길 수 있다는 단점을 가진다. 이러한 단점을 개선하기 위해서 김병기[5]는 지역적 칼라 양자화를 제안하였고 Wei-Yuan [6]는 기본 칼라 대표 값을 21개 정하고 또 다른 21개의 칼라 대표 값을 입력 영상에 따라 적응적으로 찾아내는 방법을 제안했다.

Zhong[7]은 공간적 변화도와 칼라양자화를 결합하였다. 영상을 직접 처리하는 방법으로 Zhong[7]은 공간적 변화도를 이용했고 Ohyal[10]는 지역적 이진화방법을 이용한 것이 있다. 또한

Ye[8]은 문자가 에지의 쌍을 가지고 있다는 특성을 모델링하여 은행수표에서 문자를 추출하는데 응용했으며 Gu[9]은 자연영상에서 문자열을 추출하기 위해 차분 톱햇(Differential top-hats)연산을 제안했다. 문자획의 크기에 따라 8단계로 나누고, 하얀 문자와 까만문자를 나누어 제안한 방법을 적용하여 견고하게 문자열을 추출하고 있지만 많은 수행시간이 예상된다.

본 논문에서는 모폴로지 연산인 WTH(white top-hats)과 BHT(black top-hats)를 사용하여 문자열의 대략적인 위치를 찾아낸다. 각 문자의 크기에 따라 단계적으로 적용하거나 칼라의 대표 값에 따라 각 칼라공간에 대해 문자추출과정을 반복 수행하여 많은 계산비용이 소모되지 않도록 영상을 직접 처리하였다. 영상을 직접 처리하는 방법으로 공간적 변화도가 높은 영역을 찾는 방법은 문자열 후보영역을 찾기 위한 추가적인 과정이 필요하다. WTH+BTH연산은 문자의 대략적인 위치와 영역을 동시에 찾아준다는 점에서 잇점이 있다고 하겠다.

3. 자연영상에서 문자영역 추출의 설계

자연영상에서 문자열을 추출하는 과정은 세 단계를 거친다. 첫번째는 영역화(Segmentation) 단계이고 두번째는 영역분류 단계(Block Classification)이고 세번째는 후처리(Post Processing)이다.

3.1 영역화(Segmentation) 과정

영역화 과정에서 사용한 WTH+BTH에 대해 간략히 설명하면 다음과 같다. 영상 f 에 대한 WTH(white top - hats)은 원영상 f 와 원영상에 대한 opening 연산 γ 의 결과와의 차이이다.[12]

$$WTH(f) = f - \gamma(f)$$

구조요소보다 작은 피크(peak)점들이 추출된다. 영상 f 에 대한 BHT(black top - hats)은 원영상의 closing 연산 ϕ 의 결과에 원영상을 산술적으로 뺀것이다.

$$BTH(f) = \phi(f) - f$$

구조요소보다 작은 골(trough)점들이 추출된다. WTH+BTH은 영상의 상대적 대비에 무관하게 구조요소를 포함할 수 없는 모든 영역-즉 피크(peak)와 골(trough)-를 추출하고 식은 아래와 같다.

$$\rho = WTH + BTH = f - \gamma + \phi - f = \phi - \gamma$$

문자열은 피크(peak)점과 골(trough)가 동시에 존재하므로 위 연산을 문자열에 적용하면 하나의 영역을 형성시킨다. 아래 그림은 본논문에서 적용한 연산의 결과와 공간적 변화도가 큰 부분을 찾아낸 결과를 비교한 그림이다.

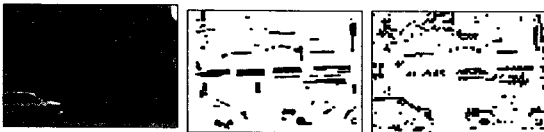


Fig.1. 원영상

Fig.2. WTH+BTH를 적용한 경우

Fig.3. 공간적 변화도 적용

위 그림에서처럼 WTH+BTH를 적용할 경우 좀더 명확한 영역을 얻을 수 있었다.

영역화 단계에서는 RGB칼라 공간에서 WTH+BTH를 수행한다. 문자열이 존재하는 곳에서는 보통 주위보다 칼라의 변화도가 상대적으로 크다는 가정하에서 WTH +BTH 연산을 수행하여 $T_{seg}(=40)$ 값보다 큰 영역을 표시하면 문자열의 위치를 찾아낼수 있다. 모폴로지 연산을 수행하기 위해서는 적당한 구조요소(structuring element)를 모델링해야 한다. 톱햇(top-hats) 연산을 적용하면 구조요소보다 작은 부분을 찾을 수 있기 때문에 찾고자 하는 문자의 최대크기에 제한을 둬으로써 구조요소를 정할수 있다. 따라서 본 논문에서는 문자획의 굵기를 25픽셀이하로 제한하여 구조요소를 25x1로 실험하였다. $T_{bigstroke}$ 보다 큰 획을 제거한다. 본논문에서 사용하는 입력영상 50개를 대상으로 사람의 눈으로 확실히 식별가능한 가장 작은 문자의 크기가 대략 7픽셀의 높이를 가지므로 7x7영역에서 밀도가 높은 영역을 찾아내어 영역화의 결과로 처리했다.

3.2 영역분류화(Block Classification)

영역분류화 단계는 문자열후보영역에 대해서 문자열-비문자열 판별을 수행한다. 영역화(Segmentation)의 결과로 추출된 영역에서 레이블링을 수행하여 각 영역을 포함하고 2픽셀 확장한 최소문자열 후보영역을 잡아준다. 영역분류화 단계는 휴리스틱을 적용하여 진행된다. 기존의 연구에서 문자열 분류화 과정에서 적용한 특성치 중에서 일부를 본 논문에서 사용했다. 최소 문자열 후보영역에 대해 Otsu알고리즘을 적용하여 이진화 한 후에 다음과 같은 조건을 검사한다. 임계치는 영역화에서 추출된 최소후보영역에서 표본 60개를 선정하여 정했다.

1. 일반적인 특성

너비 > T_{width} , 높이 > T_{height} , 면적 > T_{area} 를 만족해야 한다.

2. 전체 픽셀 수에 대한 화이트픽셀 비율 > T_{wratio}

전체 픽셀 수에 대한 블랙픽셀 비율 > T_{bratio}

3. 전이도(Transition)가 임계치를 만족해야 한다.

화이트에서 블랙으로 그리고 그 역으로의 전이가 발생되는 횟수를 Trans라고 한다면

$$Trans / (너비 + 높이) > T_{trans}$$

4. 에지의 방향이 0도와 90도에 많이 분포해야 한다.

Sobel 마스크연산(Sobel mask operation)을 적용하여 에지의 각도 θ 와 변화의 크기 G 를 계산한다. i 가 0이면 $0 \sim 22.5^\circ$ i 가 1이며 $22.5 \sim 45^\circ$...의 각도를 나타내도록 아래와 같이 i 를 계산하고 8개의 각도에서 변화의 크기 u_i 를 아래와 같이 계산한다.[6]

$$i = \text{mod}(\text{round}(\theta/22.5) + 8, 8)$$

$$u_i = u_i + G \quad (i=0, 1, \dots, 7)$$

$$TextMode = \frac{(u_0 - u_1) + (u_0 - u_2) + (u_0 - u_3)}{u_{avg}} + \frac{(u_4 - u_5) + (u_4 - u_6) + (u_4 - u_7)}{u_{avg}}$$

u_{avg} 는 $u_i (i=0, 1, \dots, 7)$ 의 평균이다. TextMode의 값이 크면 text block이다. 그러나 창틀이나 우연히 존재하는 수직, 수평에지가 자연영상에는 많이 존재하므로 TextMode가 T_{edge} 보다 작은 것을 비문자열로 분류하였다.

5 밀도변량이 낮아야 한다.

어떤 블록이 문자열을 포함하고 있다면 전경색이 고루게 분포되는 경향이 크다. 영역을 8개의 블록으로 나누어 각 영역에 대한 밀도를 $D_i(i=0,1,\dots,7)$ 이라고 하면 밀도변량 DV를 다음과 같이 정의한다.[6]

$$DV = \frac{\sum_i |D_i - D_{avg}|}{D_{avg}}$$

D_{avg} 는 D_i 의 평균을 의미한다. DV가 작으면 문자열일 가능성이 크다

6. 홀이 적고 연결요소의 크기가 유사한 것들이 일정한 간격으로 존재하면 문자열이다.

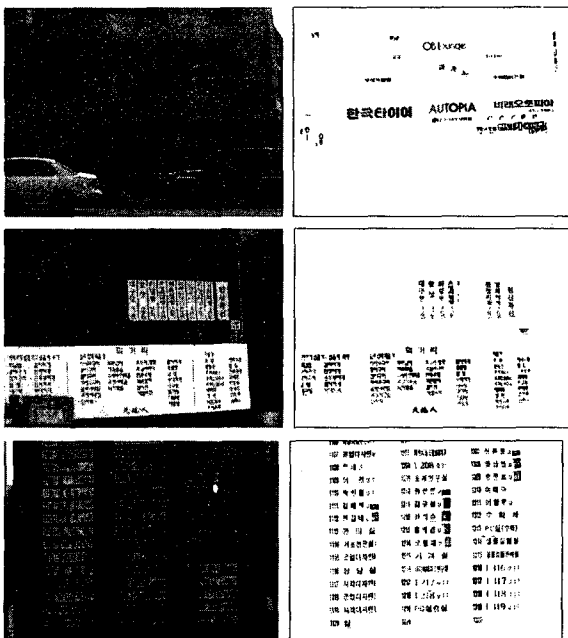
3.3 후처리

영역내에서 주요 연결컴포넌트와 유사한 위치,크기를 가진 것을 인접영역에서 찾아 결합하여 문자열의 좀더 완전한 형태를 보여준다.

4. 실험 및 결과

본 논문에서는 실험은 펜티엄 셀러론 300a PC에서 Visual C++ 6.0을 사용하여 구현하였다. 실험영상을 디지털카메라로 자연영상에서 직접 획득하였으며 한 장의 크기는 764x573 칼라영상이다.

문자열의 위치는 비교적 잘 찾아내지만 문자 하나만을 포함한 경우에는 영역분류화과정에서 비문자열로 분류하는 경우가 많다. 한글의 경우는 수직, 수평에지들이 많이 분포하므로 문자하나인 경우에도 잘 찾아진다.



문자열 추출을 위한 WTH+BTH를 적용하여 문자열의 위치를 비교적 명확히 찾아내었다. 영역분류화에서 여러문자들이 배열되어 있는 문자열에 대해서는 문자열로 처리해준다.

5. 결론

본 논문은 복잡한 자연영상에서 문자열을 추출하기 위해 WTH+BTH를 영역화에 적용했다. 영역화의 결과를 보면 기울어진 문자열에 대해서는 평행사변형모양으로 문자영역을 찾아낸다. 하지만 기울어진 영역을 처리하는 과정에서 기울어진 영역을 그것을 포함하는 최소직사각형으로 레이블링함으로써 기울어짐 정보를 없애게 된다. 이러한 단점으로 인해 기울어진 영역에 대한 영역처리를 개선점으로 남긴다. 영상이 너무 밝은 경우에는 영역화로 인해 찾아낸 영역이 상대적으로 크고 많아지므로 톱햇(top-hats)연산의 임계치를 적용적으로 결정함으로써 본 논문을 개선하고자 연구하고 있다. 본논문에서는 영문자 하나만을 포함할 경우 비문자열로 처리하는 경우가 많다. 이러한 문제등을 개선하고 추출을 견고하게 하기 위해서는 빠른 문자 인식시스템과 함께 개발되어야 한다고 본다.

6. 참고문헌

- [1]. 김희승, 영상인식, 생능출판사, 서울, 1993.
- [2]. E.G. Johnston, Printed text discrimination, Comp. Graphics Image Process 3(1) (1974) 83-89.
- [3]. K.Y. Wong, R.G. Casey, and F.M. Wahl, "Document analysis system", IBM J. Res. Develop., vol. 6, pp.642-656, November 1982.
- [4]. G. Nagy and S. Seth, "Hierarchical representation of optically scanned documents", Proc. of 7th Int. Conf. on Pattern Recognition, pp. 347-349, 1984.
- [5]. Pyeoung-Kee Kim, "Automatic Text location in complex color images using local color quantization", Proceedings of the IEEE Region 10 Conference, Vol.1, p629 -632, 1999.
- [6]. Wei-Yuan Chen, Shu-Yuan Chen, "Adaptive page segmentation for color technical journals' cover images", Image and Vision Computing, vol.16, p855-877, 1998.
- [7]. Y. Zhong, K. Karu, A.K. Jain, "Locating text in complex color image", Pattern Recog, vol.28, p1528-1535, 1995.
- [8]. Xiangyun Ye, Mohamed Cheriet, Ching Y. Suen, "Model-Based Character Extraction form complex Backgrounds", Proc. of the Fifth Int. Conf. on Document Analysis and Recognition, p511 - 514, 1999.
- [9]. Lixu Gu, Toyahisa Kaneko, "Robust extraction of Characters from color scene Image Using Mathematical morphology", Proc. of 7th Int. Conf. on Pattern Recognition, Vol.2, p1002-1004, 1998
- [10]. Jun Ohya, Akio Shio, Shigeru Akamatsu, "Recognizing Characters in Scene Images", IEEE PAMI, Vol.16, NO.2, February 1994
- [11]. Jiangying Zhou, Danial Lopresti, "Extracting text from www Images", Proc. of Int. Conf. on Document Analysis and Recognition, Vol.1, p248 -252, 1997.
- [12]. P.Soiile, Morphological Image Analysis, Springer, 1999.