한국컴퓨터정보학회 하계학술대회 논문집 제18권 제2호 (2010. 7)

캡션 내 문자와 배경의 명암값 결정

안권재°, 김계영**
°숭실대학교 컴퓨터학과
**숭실대학교 컴퓨터학부
e-mail: ankwonjae@naver.com, gykim11@ssu.ac.kr

Determing intensity value of characters and backgrounds on caption

Kwon-Jae An°, Gye-Young Kim**
°Dept. of Computing, Soongsil University
**School of Computing, Soongsil University

● 요 약 ●

본 논문에서는 동영상에서 비교적 단일 색상의 배경과 문자를 갖는 캡션을 문자인식을 위하여 문자와 배경간의 명암값 결정에 관한 내용이다. 먼저 캡션에 대해 그레이 스케일로 전환을 한 후, Otsu 방법[1]을 이용하여 이진화를 수행한다. 이 후 이진화 영상에서 흰색영역 검은색영역에 대해 각각 최대 내접 정사각형을 산출한다. 다음으로 각각의 영역에서 산출된 최대 내접 정사각형의 분산의 대소를 비교하여 문자영역과 배경영역을 결정한다. 이후 전역적인 잡음을 제거하기 문자영역에 대해 Otsu 방법을 이용하여 최종 문자영역을 결정한다. 제안된 방법의 문자영역의 명암값 결정 정확도는 약 99%로 매우 우수한 성능을 보였다.

키워드: 문자 추출(Text segmenatation), 문자 인식(Character recognition), 최대 내접 정사각형(Maximum Inscription Squares)

I. 서론

멀티미디어 시대로 접어들면서 각종 멀티미디어의 콘텐츠에 대한 수요가 급격히 증가하고 있다. 특히 멀티미디어 콘텐츠 중 동영상은 음악, 문자, 영상 등이 포함되어 있는 대표적인 멀티미디어의 콘텐츠이다. 다량의 동영상이 매일 제작이 되고 있으며, 이러한 동영상에 대해 검색을 위해 색인을 하는 작업이 이뤄지고 있다. 하지만 이러한 색인을 수작업으로 할 경우 주제에 맞는 색인에 대한 정확도는 보장 되지만, 많은 노동력이 든다는 단점이 있다. 자동으로 색인을 할 경우 주제에 맞는 색인이 힘들기 때문에 동영상 내에 있는 캡션을 찾아 캡션 내에 문자를 인식하여 해당 정보를 이용하여 색인을 한다. 이러한 경우 노동력이 매우 적게 들지만, 캡션탐지와 문자인식률에 따라 정확도가 많이 달라진다.

동영상에서 추출된 캡션 내 문자인식을 위한 전처리는 문서상의 문자인식의 전처리와 비교 할 때 문자영역추출이라는 전처리가 추가가 된다. 문서에서 문자영역과 배경영역은 각각 검은색과 흰색으로 뚜렷히 정해져 있기 때문에 간단한 이진화로 문자영역과 배경역역을 나누어 얻을 수 있는 반면, 캡션영역의 배경과 문자는 색상은 캡션 마다 매우 상이하기 때문에 이진화를 수행하여도 어떠한 영역이 문자영역인지 배경영역인지 결정할 수가 없다. 이러한 경우 휴리스틱한 방법으로 이진화된 영상에서 더 많은 화소를 갖는 부류가 배경영역으로 정할 수도 있고, 영상의 테두리에 많이 분포한 부류를 배경영역으로 정하는 직감적인 방법 등을 쓸 수 있으나, 배경이 문자영역보다 작을 경우와 문자영역이 테두리에 많이 겹칠 경우 정확도가 떨어진다. 본 논문은 추출된 캡션영역에 대해 문자인식을 위하여 문자는 배경에 비해 문자의 두께는 비교적 고르다는 특성을 이용하여 문자영역과 배경영역에 대해 명암값을 정하는 방법에 관한 것이다.

본 논문에서 2장은 관련연구 중 대표적 연구에 관한 내용이며, 3장은 제안된 내용에 관하여 자세히 설명한다.

II. 관련 연구

1. 관련연구

[2]은 Song등이 제안한 캡션영역에서 문자영역을 추출하는 대표적인 방법이다. Song의 방법은 해당 논문에서 제안된 2개의 3X3 에지 추출 마스크(그림 1)와 식(1)을 이용하여 흰색에지와 검은색에지를 구한 후 이들의 비율을 구하고 외곽 에지를 제거한 후 다시 흰색에지와 검은색에지의 비율을 구하여 이를 이용하여 문자와 배경영역의 명암 관계를 분류하는 방법이다.

$K_b(x,y)$

0	1	0
1	-4	1
0	1	0

$K_w(x,y)$

0	-1	0
-1	4	-1
0	-1	0

그림 1. [6]에서 제안된 에지 검출 마스크

$$P(x,y)=\begin{cases} WhieEdge\,,\ K_w(x,y)>0 \\ BlackEdge\,,\ K_b(x,y)>0 \\ NonEdge,\ K_w(x,y)\le 0 \\ \qquad\qquad \text{and} K_b(x,y)\le 0 \end{cases} \quad (1)$$

해당 방법은 매우 빠르며 정확도 또한 약 98.5%로 매우 높다. 하지만 이 방법은 배경영역이 확보가 되지 않아 문장영역이 영상의 테두리에 닿아 있거나, 배경영역과 문자영역의 명암대비가 매우 적을 경우 잘못된 결과가 나올 수 있다.

III. 본론

3.1 이진화

최초 캡션영역을 그레이스케일로 변환을 한다. 영역을 분류하기 위해 이진화를 수행하게 되는데 이는 문자영역과 배경영역을 분리하기 위함으로서 이진화 방법에는 여러 가지가 있으나, 본 논문에서는 Otsu의 방법을 이용하여 임계값 T_1을 구하고 식(2)에 따라 이진화를 수행하였다.

$$f(x,y)=\begin{cases} \text{if}\ \ I(x,y)\ge T_1 & 255 \\ \text{if}\ \ I(x,y)< T_1 & 0 \end{cases} \quad (2)$$

$I(x,y)$는 영상에서 해당 좌표 x, y에 대한 명암값이고 $f(x,y)$는 이진화 된 영상이다.

3.2 최대 내접 정사각형 산출

문자들은 배경에 비해 두께가 비교적 고르다. 이러한 특성을 이용하여 배경영역과 문자영역에 대한 명암값을 결정하게 되는데 이진화된 영상에서 이러한 정보를 얻기 위해 최대 내접 정사각형을 이용한다.

최대 내접 정사각형은 적응적 형태학연산을 이용하여 구할 수 있다. 영상을 읽어가면서 사각형을 확장시켜 나가며 식(3)에 따라 최대 내접 정사각형의 영역정보를 산출한다.

다음은 최대 내접 정사각형을 산출하는 자세한 과정이다.

1. 최초 이진화 영상에서 식(2)에 따라 각각의 맵을 만든다.

$$Map_{255}(x,y)=\begin{cases} \text{if}\ \ f(x,y)=255 & TRUE \\ \text{if}\ \ f(x,y)=0 & FALSE \end{cases} \quad (3)$$

$$Map_0(x,y)=\begin{cases} \text{if}\ \ f(x,y)=0 & TRUE \\ \text{if}\ \ f(x,y)=0 & FALSE \end{cases}$$

2. 각각의 맵을 읽어가며 식(4)에 만족할 때까지 반복적으로 정사각형을 늘린다.

$$\begin{aligned} &\text{if}\ \ Map_q(x+k,y+k)= TRUE \\ &\text{and}\ \ Map_q(x+k-m,y+k)= TRUE \\ &\text{and}\ \ Map_q(x+k,y+k-n)= TRUE \\ &then\ \ k=k+1 \\ &where\ 0\le m,n\le k,\ k\text{의 초기값은}\ 0 \end{aligned} \quad (4)$$

3. 만약 식(5)를 만족하면, i번째 최대 내접 정사각형이 얻어진다.

$$\begin{aligned} &\text{if}\ \ Map_q(x+k,y+k)= FALSE \\ &\text{and}\ \ Map_q(x+k-m,y+k)= FALSE \\ &\text{and}\ \ Map_q(x+k,y+k-n)= FALSE \\ &then\ \ IS_i^q=k \\ &where\ 0\le m,n\le k \end{aligned} \quad (5)$$

여기서 Map은 TRUE, FALSE로 구분되는 맵이고, 밑 첨자 q는 맵 및 이에 대응하는 최대 내접 정사각형의 식별자이고 IS는 최대 내접 정사각형이다.

3.3 화소색 결정

앞서 얻는 최대 내접 정사각형의 넓이의 값을 이용하여 영역의 고른 정도를 분별한다. 이를 위해 최대 내접 정사각형의 평균 및 분산을 식(6), 식(7)에 따라 구하고 분산을 이용하여 고른 정도를 분별하여 문자는 배경에 비해 비교적 너비가 고르다는 특성을 이용하여 분산이 작은 영역을 문자영역, 분산이 큰 영역을 배경영역으로 지정한다.

$$E(IS^q)=\frac{1}{N_q}\sum_{i=1}^{N_q} IS_i^q \quad (6)$$

$$\sigma_q^2=\frac{1}{N_q}\sum_{i=1}^{N_q}\left\{E(IS^q)-IS_i^q\right\}^2 \quad (7)$$

여기서 E는 평균이고 N_q는 해당 맵에서 얻어진 최대 내접 정사각형의 개수이고, σ^2은 그에 분산이다.

3.4 최종 화소색 결정 및 잡음 제거

캡션의 특성 상 비교적 단색의 배경 및 문자라해도 원배경과의 오버레이에 의해 전역적 이진화만을 수행 하였을 때 잡음이 심하다. 본 논문에서는 이러한 잡음을 제거하기 위해 앞서 얻어진 문자영역에서만 Otsu 방법을 수행하여 오버레이 등으로 인한 잡음을 제거한다. 이 방법을 적용하기 전 우선 캡션 내 배경영역과 문자영역의 명암값 관계를 고려해야한다. 배경영역과 문자영역은 관계는

[2]에서는 4가지의 경우, 배경영역과 문자영역이 어두운 경우(DonD), 배경영역과 문자영역이 밝은 경우(WonW), 배경영역은 어둡고 문자영역이 밝은 경우(WonD), 배경영역은 밝고 문자영역은 어두운 경우(DonW)로 구분하고 있으나 본 논문에서는 문자영역이 배경영역보다 상대적으로 밝은 경우와 문자영역이 배경영역보다 상대적으로 어두운 경우, 두 가지로 구분한다.

앞서 얻어진 문자영역이 배경영역보다 상대적으로 밝은 경우면 최초 이진화 단계에서 얻어진 임계값 T_1부터 255사이의 명암값들을 이용하여 새로운 임계값 T_2를 구하고 T_2보다 값이 큰 화소들을 최종문자영역으로 정하고 해당 화소들은 검은색 0, 이외의 화소들은 흰색 255로 정한다. 반대로 문자영역이 배경영역보다 상대적으로 어두운 경우면 0에서 T_1사이의 명암값들을 이용하여 새로운 임계값 T_2를 구하고 T_2보다 값이 작은 화소들을 최종문자영역으로 정하고 해당 화소들은 검은색 0, 이외의 화소들은 흰색 255로 정한다.

IV. 실험결과

실험에 쓰인 캡션은 뉴스에서 얻었으며, 본 논문은 문자영역과 배경영역의 화소색 결정을 위한 논문이기 때문에 캡션을 자동으로 검출 한 것이 아니며, 수작업으로 250개를 모았다.

그림 2.는 원영상 및 Otsu 방법을 이용한 이진화 결과와 화소색이 결정, 잡음이 제거 된 영상을 보여준다.

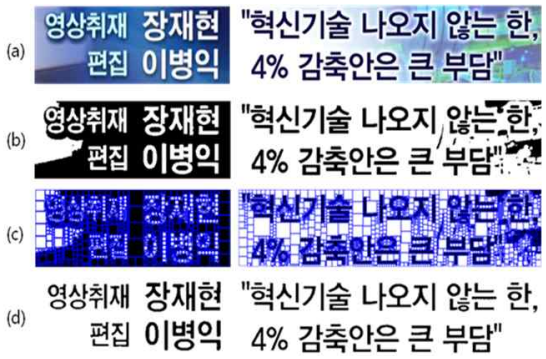

그림 2. (a)는 원영상, (b)는 Otsu 방법을 이용한 이진화 영상, (c)는 최대 내접 정사각형을 구하는 과정, (d)는 최종 문자영역 및 배경영역의 화소색이 결정된 영상

본 논문에서 제안된 방법으로 문자영역 및 배경영역의 화소색 결정하였을 때 표 2에서 보는 것과 같이 정확도는 약 99%를 보였다.

표 1. 본 논문에서 제안한 방법에 의한 문자영역 및 배경영역의 화소색 결정에 관한 결과

실험에 쓰인 캡션 수	정확한 결과의 수 (문자영역 검은색, 배경영역 흰색)	정확도(%)
250	249	99.6%

잡음 제거 과정을 수행 한 후, 문자인식을 수행 하였을 때 문자인식률은 상승하였다. 문자인식은 상업용 문자인식 소프트웨어 및 구현한 문자인식 프로그램을 사용하였으며, 결과는 표 2와 같다.

문자인식기는 [3]의 방법인 인공신경망을 이용하여 만들었으며 문자 간 접촉현상 해결은 [4]의 방법에 기반하여 만들었다.

표 2. 문자인식 결과

실험에 쓰인 캡션 수	잡음제거 전 문자 인식률(%)	잡음제거 후 문자 인식률(%)
상업용 문자인식기	94	96
[3][4]번의 방법으로 구현한 문자인식기	87	91

구현한 문자인식 프로그램이 인식률이 해당 논문에 나온 것보다 낮은 이유는 학습데이터의 양이 적고 인공신경망의 구조를 최적화하지 않았기 때문인 것으로 추측된다.

V. 결론

본 논문은 동영상에서 얻어진 캡션에서 문자인식을 위한 문자영역과 배경영역에 대한 화소색 결정에 관한 방법을 제안했다. 문자영역과 배경영역의 명암값 관계에 대한 사전정보 없이 문자영역의 화소색은 검은색 배경영역에 화소색은 흰색으로 정할 수 있으며, 오버레이에 의해 이진화 결과가 잡음이 심하더라도 비교적 단색의 배경 및 문자의 캡션에서 정확도는 약 99%를 보였다.

본 논문에서는 최초 이진화를 수행하여 2부류에서 문자영역을 결정하였다. 앞으로 적절한 색상 양자화 연구를 통하여 다부류에서 문자영역 결정하는 방법을 연구할 계획이다.

참고문헌

[1] N. Otsu., “A threshold selection method from gray-level histograms”, IEEE Transactions on Systems, Man and Cybernetics, vol.SMC-9, pp. 62-66, 1978년

[2] Jiqiang Song, Min Cai, Michael R.Lyu, “A ROBUST STATISTIC METHOD FOR CLASSIFYING COLOR POLARITY OF VIDEO TEXT”, 2003. ICME ’03. Proceedings. 2003 International Conference on Volume 2, pp. 581-584, 2003년

[3] Jin-Soo Lee, Oh-Jun Kwon, Sung-Yang Bang, “Highly accurate recognition of printed Korean characters through an impoved two-stage classification method”, Pattern Recognition vol.3, pp.1935-1945, 2003년

[4] 김의정, “칼라 문서에서 문자 영역 추출 및 문자 분리”, 한국퍼지 및 지능시스템학회 논문지, Vol.9, No.4 pp.444-450, 1999년