

스트록 필터를 이용한 문자영역 이진화에 관한 연구

정회원 정철곤*, 김중규*

A Study on Automatic Binarization of Text Region Using a Stroke Filter

Cheolkon Jung*, Joongkyu Kim* *Regular Members*

요 약

비디오 문자는 중요한 내용정보를 담고 있기 때문에 비디오의 내용 해석에 있어서 매우 중요한 정보이다. 본 논문에서는 스트록 필터를 이용해 자동으로 문자영역을 이진화하는 방법을 제안하였다. 제안된 문자 이진화 방법은 스트록 필터에 의한 문자컬러극성 결정단계, 스트록 필터의 응답치에 대한 이진화 단계, 그리고 국소 영역 확장 단계로 구성되어 있다. 본 방법은 스트록 필터의 응답치를 이용해 문자의 컬러 극성을 결정함으로써 극성 결정 성능이 뛰어나다. 또한 문자의 획 특성을 고려해 문자영역을 이진화하기 때문에 배경영역의 변화에 대하여 강인한 이진화 성능을 나타낸다. 다양한 비디오 영상에 대하여 실험한 결과, 이진화 성능이 우수함을 확인할 수 있었다.

Key Words : Text Binarization, Color Polarity, Stroke Filter, Local Region Growing

ABSTRACT

The videotext brings important semantic clues into video content analysis. In this paper, we propose an automatic binarization method of text region using a stroke filter. Proposed text binarization method consists of stroke filtering, text color polarity determination, and local region growing. By using the responses of dark and bright stroke filters, we can determine color polarity of text region automatically. And the method is robust against complex background, because it considers stroke information of videotexts by using a stroke filter. The effectiveness of our method is verified by experiments on a challenging database.

I. 서 론

영상이나 비디오의 문자는 중요한 내용정보를 담고 있기 때문에 내용 해석에 있어서 매우 중요하다. 영상에 포함된 문자 정보를 추출하는 기술은 크게 문자검출기술, 문자 이진화 기술, 문자인식기술의 세 가지로 구성이 된다^[1-4]. 문자검출기술은 영상에 포함된 문자들의 위치를 검출하는 기술을 말하며, 문자 이진화 기술은 검출된 문자영역에서 문자를 이진화하는 기술이며, 문자인식기술은 이진화된 문

자영역을 인식하는 기술을 말한다(그림 1).

본 논문에서 제안된 방법은 문자 이진화 기술에 관한 것이다. 영상에서 문자영역을 배경으로부터 이진화하는 종래의 방법은 임계값에 의한 방법, 클러스터 기반 방법, OCR 인식에 의한 방법으로 분류된다. 임계값에 의한 방법은 고전적인 방법으로 1979년, Otsu 의해 제안된 방법이 대표적이다^[5]. 이 방법은 배경과 문자영역의 밝기 값 분포에 대한 분산차를 최대화 시키는 값을 임계값으로 선택하는 방법이다. 이 방법의 경우, 문자영역과 배경영역의

* 성균관대학교 정보통신공학부 (ckjung@ece.skku.ac.kr, jkkim@skku.edu)

논문번호 : KICS2007-08-385, 접수일자 : 2007년 8월 31일, 최종논문접수일자 : 2008년 1월 4일



그림 1. 문자정보추출기술



그림 2. 문자 이진화 오류-1



그림 3. 문자 이진화 오류-2

밝기 차이가 현저한 경우에 이상적이며, 만약 유사한 문자영역의 밝기와 비슷한 밝기를 가지는 영역이 배경에 나타날 경우, 이진화에 실패하게 된다. 1996년, Lienhart에 의해 제안된 클러스터링 방법은 컬러 값의 개수를 줄여 후보영역을 생성한 후, 크기 등의 제한 조건에 의해 영역 필터링을 하여 문자영역을 이진화한다⁶⁾. 이때 문자 영역은 유사한 컬러 값을 가진다는 가정에 의해 결정 된다. 따라서 임계값에 의한 방법과 마찬가지로 배경에 문자영역과 유사한 컬러를 가지는 영역이 나타날 경우, 이진화에 실패하게 된다. 또한 이 두 방법은 공통적으로 문자영역의 컬러에 대한 극성을 고려하지 않으므로 극성을 결정해 주는 과정이 포함되어야 한다. 또한 D. Chen에 의해 제안된 OCR 인식에 의한 방법은 여러 개의 임계값을 설정하여 문자영역을 이진화한 후, OCR을 이용해 각 영역에 대해 인식을 수행해 가장 인식결과 값이 가장 높은 경우를 문자영역 이진화 결과로 결정한다⁷⁾. 이 방법의 경우, 문자영역의 컬러 극성을 함께 연어낼 수 있는 장점이 있지만, 다양한 경우에 대하여 문자인식을 수행하기 때문에 처리시간이 많이 걸리는 단점이 있다.

그림 2는 배경영역에 문자영역의 밝기와 유사한 영역이 존재할 경우에, 기존의 방법으로 문자영역을 이진화할 경우 발생하는 문제점에 대한 예시이다. 그림에서 보는 바와 같이 배경영역이 포함되어 문

자영역으로 이진화된다. 이는 배경영역과 문자영역의 밝기 값이 유사하기 때문에 나타나는 현상이다. 이러한 오류는 문자의 높이, 너비, 높이와 너비 비율 등의 휴리스틱한 정보에 의해 어느 정도 제거가 가능하지만, 한계가 있다.

그림 3은 컬러 극성을 잘못 결정하여 발생하는 문자이진화 오류이다. 컬러 극성을 어둡음으로 결정하였기 때문에 발생하는 오류이다. 그림에서 보는 바와 같이 컬러 극성을 잘못 결정해 준 경우, 문자영역을 제대로 이진화할 수 없게 된다.

본 논문에서는 스트록 필터를 이용해 이러한 두 가지 오류를 해결하고, 정확한 문자영역을 이진화하는 방법을 제안하였다. 스트록 필터에 의한 응답값을 이용해 컬러 극성을 결정하였고, 스트록 필터를 이용해 스트록 특성을 가지지 않는 배경 영역을 제거하였다.

II. 제안된 알고리즘

2.1 알고리즘 개요

문자 이진화시 발생했던 문제점인 그림 2, 3의 오류를 해결하기 위해 스트록 필터를 이용하였다. 스트록이란 문자를 구성하는 획을 의미하는데, 스트록 필터는 이를 검출할 수 있는 필터를 말한다. 이러한 스트록 필터를 이용해 배경에서 발생하는 오류영역을 제공할 수 있으며, 스트록 필터의 response에 의해 컬러 극성을 효율적으로 결정할 수 있었다. 제안된 문자이진화 방법은 그림 4와 같이 크게 스트록 필터링 (II.2), 문자 컬러 극성 결정 (II.3), 국소영역확장 (II.4)로 구성된다. 여기서 입력 영상은 문자검출 과정에 의해 생성된 text sub-image이며, 출력 결과는 문자영역 이진화 결과이다.

2.2 스트록 필터링

스트록 필터는 영상에서 스트록을 검출하기 위한 필터이며, 국부적인 영역 해석에 기반한 필터이다. 그림 5에서 보는 바와 같이, 스트록 필터는 중심픽셀 (x, y) 의 주위로 3개의 직사각형 영역으로 구성된다. 그림에서 1, 2, 3은 각 영역의 번호를 의미하며, a 는 방향을, d 는 세로 너비를, w 는 가로너비를 의미한다. 상반된 두 가지 컬러 극성을 모두 고려하기 위해 아래와 같이 bright 와 dark 스트록 필터의 응답값 R_B 와 R_D 를 사용하였다.

$$R_B(a,d) = m_1 - m_2 + m_1 - m_3 - |m_2 - m_3| \quad (1)$$

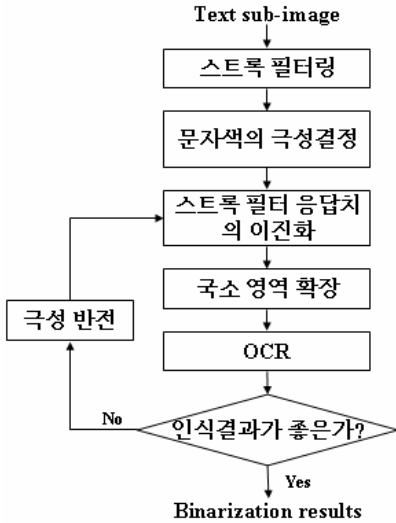


그림 4. 알고리즘의 흐름도

$$R_D(a,d) = m_1 - m_2 + m_1 - m_3 - |m_2 - m_3| \quad (2)$$

여기서 m_k 은 k 번째 영역의 평균을 의미한다. 또한 integral image 를 사용하여 스트록 필터링의 계산량을 최소화 하였다^[8].

2.3 컬러 극성 결정

II.2 절의 스트록 필터의 응답 값을 이용해 문자의 컬러 극성을 결정한다. 이때 사용되는 measure 는 두 가지로서 다음과 같다. 첫 번째 measure 는 bright 와 dark 스트록 필터 응답치의 비율(F_R)이며, 아래 식과 같이 표현된다.

$$F_R = \sum R_B / \sum R_D \quad (3)$$

일반적으로, F_R 이 크다는 것은 문자영역이 밝은 컬러 극성을 가지고 있음을 의미한다. 하지만 F_R 이 크다는 것이 항상 문자 컬러 극성이 밝다는 것을 의미하는 것은 아니다. 예를 들어, 밝은 컬러를 가지는 배경에 밝은 컬러를 가지는 문자인 경우, 혹은 어두운 컬러를 가지는 배경에 어두운 컬러를 가지는 문자인 경우에는 성립하지 않는다. 따라서 이러한 특성을 고려한 또 다른 measure F_E 를 사용한다. F_E measure는 다음과 같이 정의된다.

$$F_E = \sum N_B / \sum N_D \quad (4)$$

여기서 N 은 스트록 필터의 응답치에 대해 이진화한

표 1. FR과 FE값의 분포

	BonD	DonB	BonB	DonD
F_R	$\gg 1$	$\ll 1$	1	1
F_E	1	1	> 1	< 1

영상에서 수평라인을 그었을 경우, black 과 white 가 서로 교차하는 교차점들의 개수이다. 즉 N_B 는 bright 스트록 필터 응답치의 이진화 결과에 대한 교차점의 개수를 말하고, N_D 는 dark 스트록 필터 응답치의 이진화 결과에 대한 교차점의 개수이다. 문자 컬러 극성과 배경 컬러 극성에 대한 이러한 두 가지 measure 값의 분포는 표 1과 같다.

여기서, BonD란 Dark 한 배경에서 Bright한 문자가 존재하는 영상을 말한다. 마찬가지로 DonB 란 Bright한 배경에서 Dark 한 문자, BonB 란 Bright한 배경에서 Bright한 문자, DonD 란 Dark한 배경에서 Dark 한 문자가 존재하는 영상을 말한다. 표에서 보는 바와 같이 F_R 이 현저하게 크거나, 작은 경우는 문자의 컬러 극성을 명백히 결정함을 알 수 있다. 하지만, BonB 나 DonD 의 경우, 모두 F_R 이 1에 가까우며 컬러 극성을 결정하는데 변별력을 갖지 못한다. 하지만, 이 경우에 FE 는 1을 기준으로BonB 의 경우 크게, DonD의 경우 작게 나타난다. 이를 이용해, BonB 나 DonD에서도 문자의 컬러 극성을 결정할 수 있게 되는데, 컬러 극성을 결정하는 방법은 다음 식과 같다.

$$\begin{aligned} & \text{if } F_R > 1.1, \text{ then ColorPolarity} = \text{Bright} \\ & \text{if } F_R < 0.9, \text{ then ColorPolarity} = \text{Dark} \\ & \text{if } 0.9 \leq F_R \leq 1.1 \text{ and if } F_E \geq 1, \text{ then ColorPolarity} = \text{Bright} \\ & \text{if } 0.9 \leq F_R \leq 1.1 \text{ and if } F_E < 1, \text{ then ColorPolarity} = \text{Dark} \end{aligned} \quad (5)$$

스트록 필터에 의한 응답치에 대한 이진화는 지정된 임계값에 의해 결정하였으며, 이진화된 영상은 영역확장을 수행하기 위한 초기 이진화 결과로 사용된다.

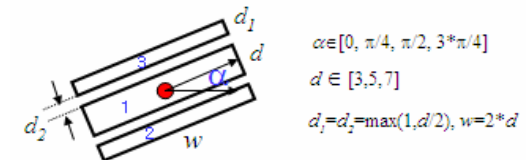


그림 5. 스트록 필터

2.4 국소 영역 확장

국소영역확장은 이진화 영상을 seed 로 하여 영역확장을 수행한다. 이때, 영역을 확장하기 위해 세가지의 제한 조건을 두었다. 국소영역확장의 구체적인 과정은 그림 6에서 기술하였다. 그림에서와 같이 먼저 이진화 된 스트로크 영상과 원 영상의 문자 영역으로부터 영역 intensity의 확률 밀도 함수 (PDF)를 구한다. 3x3 국소 윈도우를 형성하여 이진화 영상에 대하여 스캐닝을 수행한다. 이때 문자로 판별된 픽셀의 개수가 4~8 사이인 윈도우를 선택한다. 그리고, 윈도우 내에 비문자 영역으로 판별된 픽셀의 확률이 T_1 보다 크고, 이웃 텍스트 픽셀과의 intensity 차이가 T_2 보다 작은 경우에 영역 확장을 수행한다. 여기서 픽셀 intensity s 의 확률 값은 다음 식에 의해 결정된다.

$$Pr(s) = PDF(s) \quad (6)$$

문자영역 intensity의 확률밀도함수를 통해 얻은 픽셀의 확률을 제한 조건으로 두었기 때문에 보다 강건한 영역확장을 수행할 수 있다. 마지막으로, 더 이상 픽셀의 레이블이 변하지 않을 때까지 이 루틴을 반복 수행한다. 그림 7 (a)는 스트로크 필터의 응답치에 의한 초기 이진화 결과이며, 그림 7(b)는 국소영역 확장을 수행한 최종 이진화 결과이다. 마지막으로 OCR에 의해 인식을 수행한 후, 결과를 검증한다. OCR에 영역 확장된 이진화 결과를 입력하여, 인식 스코어가 높으면 최종 이진화 결과로 출력하고, 낮다면 II.3 절의 과정으로 돌아가 반대 색극성의 이진화 결과에 대해 영역확장을 수행하여 최종 이진화 결과를 출력한다.

III. 실험 결과

실험은 Intel Pentium Processor 1.40GHz 와 256 RAM에서 수행되었으며, 실험영상은 KBS, MBC, SBS로부터 획득한 435개 영상을 사용하였다. 영상의 크기는 720x480 이다. 그림 8은 본 실험에 사용된 프로그램의 사용자 인터페이스이다. 그림에서 보는 바와 같이 원영상과 문자영역 이진화 결과가 보여진다.

본 논문은 문자를 이루는 특징인 스트로크를 이용해 문자의 컬러 극성을 결정하기 때문에, 컬러 극성의 정확도가 97.4%로 매우 높다. 또한 이전의 자막 영역 검출에서 사용된 스트로크 필터의 응답치를 그

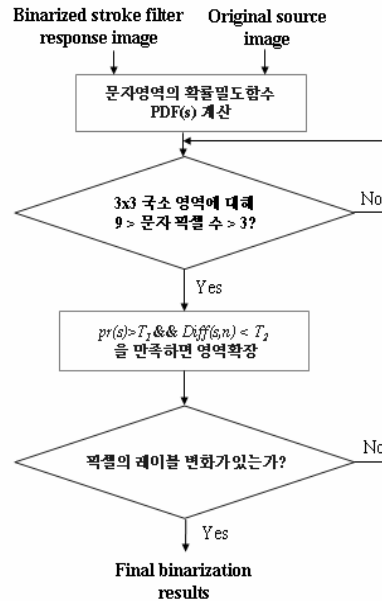


그림 6. 국소영역확장



그림 7. 국소영역확장의 예: (a) 초기 이진화 결과, (b) 국소영역확장 결과

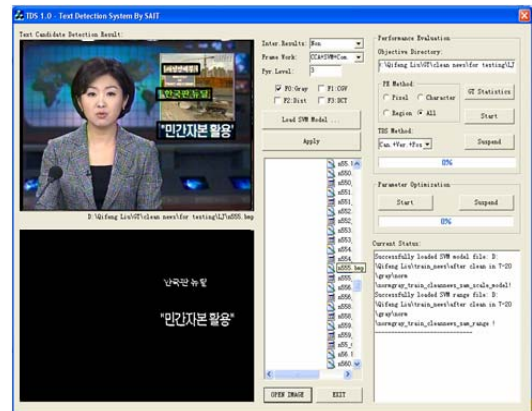


그림 8. 사용자 인터페이스

대로 사용하기 때문에 계산량을 줄일 수 있다. 뿐만 아니라, 스트로크에 의해 문자 이진화를 수행하기 때문에, 배경의 컬러 극성이 문자영역의 컬러 극성과 유사한 경우에도 강건한 특성을 나타낸다. 그림 9, 10은 이에 대한 구체적인 예를 나타낸다.

그림 9에서 보는 바와 같이 배경영역과 문자영역의 컬러 극성이 유사한 경우, 기존의 방법은 컬러

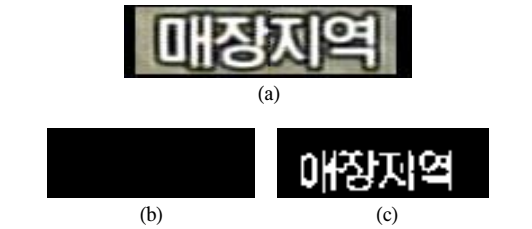


그림 9. 실험 결과-1: (a) 자막 영상, (b) 기존의 방법, (c) 제안된 방법

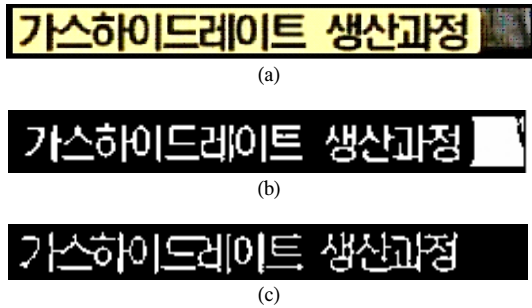
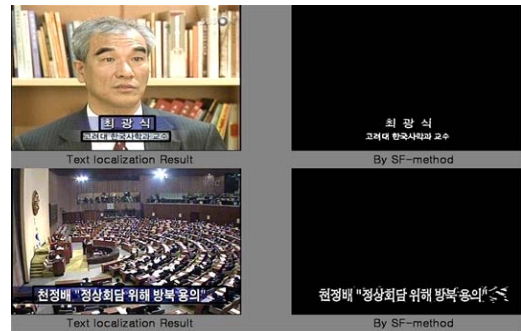


그림 10. 실험 결과-2: (a) 자막 영상, (b) 기존의 방법, (c) 제안된 방법

극성을 검출하지 못하기 때문에 문자 이진화를 수행할 수 없다. 하지만, 제안된 방법은 이러한 문제를 해결할 수 있다. 또한 그림 10과 같이 배경 영역에 문자영역과 유사한 컬러 극성을 가지는 영역이 존재할 경우 기존의 방법의 경우 제거할 수 없었지만, 제안된 방법의 경우 제거가 가능하다. 이러한 결과는 문자 인식성능에 직접적인 영향을 주기 때문에 매우 중요한 효과로 볼 수 있다. 그림 11은 또 다른 실험 영상에 대한 실험 결과이다.

그림 9에서 보는 바와 같이 배경영역과 문자영역의 컬러 극성이 유사한 경우, 기존의 방법은 컬러 극성을 검출하지 못하기 때문에 문자 이진화를 수행할 수 없다. 하지만, 제안된 방법은 이러한 문제를 해결할 수 있다. 또한 그림 10과 같이 배경 영역에 문자영역과 유사한 컬러 극성을 가지는 영역이 존재할 경우 기존의 방법의 경우 제거할 수 없었지만, 제안된 방법의 경우 제거가 가능하다. 이러한 결과는 문자 인식성능에 직접적인 영향을 주기 때문에 매우 중요한 효과로 볼 수 있다. 그림 11은 또 다른 실험 영상에 대한 실험 결과이다.

실험결과에서 보는 바와 같이 제안된 방법이 일반적인 영상에 대해 좋은 결과를 나타내지만, 다음과 같은 경우에 대해서는 단점을 가지고 있다. (1) 비디오 문자의 높이가 10 픽셀 이하인 경우와 (2) 한 text sub-image 에 2개의 상반된 문자 색 극성



(a)



(b)

그림 11. 문자 이진화 결과

을 가지는 경우에는 우수한 이진화 성능을 보장할 수 없다. 향후 연구 방향으로 이러한 문제를 해결하고자 한다.

IV. 결론

본 논문에서는 비디오 영상에서 스트록 필터를 이용해 문자영역을 이진화하는 기술을 제안하였다. 스트록 필터를 이용해 문자의 색 극성을 자동으로 결정할 수 있었다. 또한 스트록 필터의 초기 이진화 결과를 seed 영역으로 사용하여 국소영역확장을 수행하여 최종 이진화 결과를 얻을 수 있었다. 따라서 스트록 필터에 의해 문자의 획 특성을 고려함으로써 배경영역의 변화에 강인한 문자영역 이진화 성능을 얻을 수 있었다. 다양한 비디오 문자를 포함하는 비디오 영상에 대해 실험한 결과, 우수한 이진화 성능을 나타냄을 확인할 수 있었다.

참고 문헌

[1] K. Jung, K.I. Kim, A.K. Jain, "Text Information Extraction in Images and Video:

A Survey,” International Journal of Pattern Recognition, vol. 37 (5), pp. 977-997, May 2004.

[2] M. Lyu, J. Song, and M. Cai, “A comprehensive method for multilingual video text detection, localization, and extraction,” IEEE Trans. on Circuits and Systems for Video Technology, Vol. 15(2), pp. 243-255, 2005.

[3] 정철근, 김중규, “Automatic superimposed text localization from video using temporal information,” 한국통신학회 논문지, Vol. 32 (9), pp. 834 ~ 839, 2007.

[4] R. Lienhart and A. Wernicked, “Localizing and segmenting text in images and videos,” IEEE Trans. on Circuits and Systems for Video Technology, Vol. 12(4), pp. 256-268, 2002.

[5] N. Otsu, “A Threshold Selection Method from Gray-Scale Histogram,” IEEE Trans. on System Man Cybern., vol. 1, pp. 62-66, 1979.

[6] R. Lienhart, “Automatic Text Recognition in Digital Videos,” Proceeding SPIE, Image and Video Processing IV, pp. 2666-2675, 1996.

[7] D. Chen, O. Jean-Marc, and B. Herve, “Text Detection and recognition in images and video frames,” Pattern Recognition, pp. 595-608, 2004.

[8] R. Lienhart and J. Maydt. “An Extended Set of Haar-like Features for Rapid Object Detection,” IEEE Conf. Image Processing, vol. 1, pp. 900-903, 2002.

정 철 근 (Cheolkon Jung)

정회원



1995년 2월 성균관대학교 전자공학과 학사

1997년 2월 성균관대학교 전자공학과 석사

2002년 8월 성균관대학교 전기전자컴퓨터공학부 박사

2002년 10월~2007년 2월 삼성종합기술원 전문연구원

2007년 9월~현재 성균관대학교 정보통신공학부 박사후연구원

<관심분야> 멀티미디어 콘텐츠 분석, 멀티미디어 요약 및 검색, 영상처리, 컴퓨터비전, 컴퓨터그래픽스, 디지털비디오처리

김 중 규 (Joongkyu Kim)

정회원



1980년 서울대학교 전자공학과 학사

1982년 서울대학교 전자공학과 석사

1989년 The Univ. of Michigan, Ann Arbor, Department of Electrical and Computer Engineering Ph.D.

1980년~1981년 한국전자통신연구소 위촉연구원

1989년~1990년 University of Michigan, Post Doctoral Fellow

1990년~1991년 삼성전자 선임연구원

1992년~현재 성균관대학교 정보통신공학부 정교수
<관심분야> 적응신호처리, 레이더신호처리, 의학영상 신호처리, 음향신호처리, 디지털비디오처리