

MST를 이용한 문자 영역 분할 방법

진병태*, 김영인**

A Method for Character Segmentation using MST(Minimum Spanning Tree)

Byung Tae Chun*, Young In Kim**

요약

기존의 문자 영역 추출 방법은 전체 영상으로부터 컬러 영역 분할이나 프레임 차 방법을 이용하였다. 이들 방법은 휴리스틱에 많이 의존하므로 추출하려는 문자의 사전 정보를 가지고 있어야한다는 점과 구현에 많은 어려움이 존재한다. 본 논문에서는 휴리스틱한 부분을 줄이고 알고리즘을 단순화한 방법을 제안하고자 한다. 문자의 지형학적 특징점을 추출하고 이 점들을 MST(Minimum Spanning Tree)를 형성하여 문자의 후보 영역을 추출한다. 문자 영역을 후보 영역의 검증을 통하여 추출한다. 실험 결과 문자의 후보 영역 추출율은 100%이었으며 최종 문자 영역 추출율은 98.2%이었다. 또한 복잡한 영상에서 존재하는 문자 영역도 잘 추출됨을 볼 수 있다.

Abstract

Conventional caption extraction methods use the difference between frames or color segmentation methods from the whole image. Because these methods depend heavily on heuristics, we should have a priori knowledge of the captions to be extracted. Also they are difficult to implement. In this paper, we propose a method that uses little heuristic and simplified algorithm. We use topographical features of characters to extract the character points and use MST(Minimum Spanning Tree) to extract the candidate regions for captions. Character regions are determined by testing several conditions and verifying those candidate regions. Experimental results show that the candidate region extraction rate is 100%, and the character region extraction rate is 98.2%. And then we can see the results that caption area in complex images is well extracted.

▶ Keyword : 비디오 영상(video image), MST, 문자 영역 추출(character segmentation)

• 제1저자 : 진병태
• 접수일 : 2006.05.29, 심사일 : 2006.06.23, 심사완료일 : 2006.07.10
* 국립 환경대학교 웹정보공학과 교수 ** 부산대학교 바이오시스템공학부 교수

I. 서론

비디오 영상에서 중요한 정보를 추출한다면 다양한 용도로 사용할 수 있다. 비디오에서 문자 정보는 비디오 영상에서 중요한 색인 정보로 활용 될 수 있다. 예를 들면 스포츠, 제작자 이름, 장면의 위치, 방송 기자의 이름, 프로그램 소개 등은 그 의미상 중요한 정보가 될 수 있다.

기존의 문자 인식에서의 많은 접근 방법은 주로 인쇄체 또는 카메라 문서 영상 문자 인식에 중점을 두었다. 이들 시스템은 높은 정도의 인식율을 보여주고 있다[1]. 영상으로부터 문자를 추출하기 위한 기존의 연구[2,3,4,5,6]는 문자 영역 추출에 한개 또는 여러 개의 제한 사항들이 있었다. 즉, 문자의 크기, 폰트의 크기, 문자의 색상 등에 관한 제약 조건들이 존재하였다. Micheal A. Smith and Takeo Kanade는 특정 문자 정보를 가지고 있는 비디오로부터 영역 추출에 많은 연구를 진행하였다[7]. 특히 그들은 개별 문자를 분리하기 위하여나 문서 인식을 위한 분야에 많은 연구를 진행하였다.

문자 영역 추출을 위한 다른 방법은 Ohya, Shio, and Akamatsu가 제안한 방법이다[8]. 3D 공간에 존재하는 장면에서 문자를 추출하는 방법을 제안하였다. 문자의 자유도에 따라 3D 문자는 변화가 심한데 이들은 문자의 위치 및 색상을 제한하여 문자 영역을 추출하였다.

본 논문은 문자의 특징점을 추출하고 추출된 특징점을 MST 방법을 이용하여 문자 후보 영역을 추출한다. 즉, 특징점 간의 상호 관계(색상, 거리, 각도)을 고려하여 MST를 구성하여 형성된 영역을 문자의 후보 영역으로 추출한다는 것이다.

II. 문자의 특징 추출 방법

비디오 입력 영상은 그림 1과 같이 비디오 영상에 자막이 삽입된 형태를 띠고 있다. 여기서 문자 영역은 개별 문자의 구성원으로 이루어진 집합체로 볼 수 있다. 개별 문자의 명암도 특성은 그림2과 같이 배경과 문자의 명암도 차(α), 문자 획의 폭(β) 그리고 문자간 거리(γ)와 같은 3가지 특성 변수로 간략히 정의 할 수 있다. 여기서 α 값은 문자의 색상과 관련이 있고 β 값은 문자의 서체와 관련이 있다고 볼 수 있다.

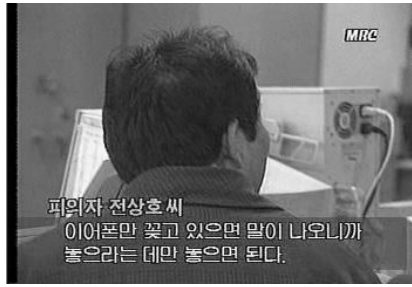


그림 1. 비디오 자막 영상
Fig.1. a text image in videos

일반화된 문자 영역 추출을 위해 앞에서 언급한 α , β , γ 값을 일반화하는 방법을 모색하여야 한다. 일반화된 α , β , γ 값을 구한다는 것은 어느 상황에도 적용 가능한 α -임계값 (Threshold value), β -임계값, γ -임계값을 구한다는 이야기와도 같다. 다시 말해서 어느 문자에 대해서도 α -임계값, β -임계값, γ -임계값이 모두 만족할 수 있는 경우를 만들겠다는 것이다.

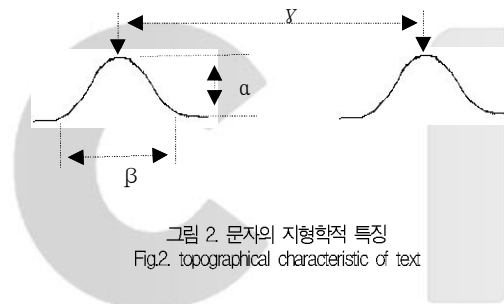
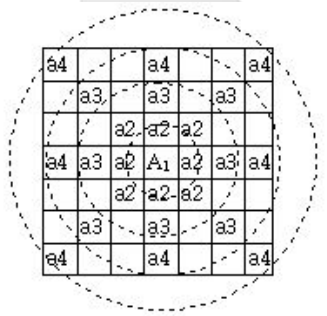


그림 2. 문자의 지형학적 특징
Fig.2. topographical characteristic of text

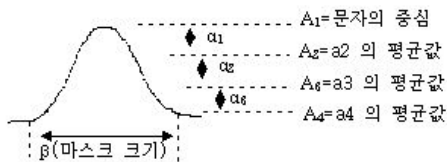
α -임계값과 β -임계값은 문자의 색상과 서체에 따라 서로 동일 영상에서 다양하게 공존할 수 있다. 예를 들어, 배경이 복잡하고 문자 색상이 여러 종류가 혼합되어있을 경우 하나의 α -임계값 결정은 문제가 발생한다. 이와 같은 다양한 문자의 특성을 만족하기 위한 α -임계값, β -임계값, γ -임계값을 어떻게 구할 것인가가 중요한 문제로 등장한다. 본 논문에서는 기존의 지형학적 문자 특징점 추출 방법[9]을 수정 보완하여 α -임계값, β -임계값을 결정하는 방법을 사용하였다.

문자의 지형학적 특징점 추출 방법은 그림3(a)와 같이 마스크 형태로 표현할 수 있다. 이 마스크에 α , β 에 관계된 정보를 관련시키면 α -임계값, β -임계값을 구할 수 있을 것으로 본다. 문자의 지형학적 특징을 보면 그림3(b)와 같이 배경 부분에서 문자 부분으로 단조 증가하다가 정점을 기점으로 감소함을 볼 수 있다. 따라서 마스크의 $A1(=a1)$ 부

분이 문자의 정점 부분에 있다고 가정하면 이웃 화소의 평균값 $A_2(3*3$ 마스킹, a_2 의 평균값), $A_3(5*5$ 마스킹, a_3 의 평균값), $A_4(7*7$ 마스킹, a_4 의 평균값),,, 은 점점 감소함을 볼 수 있을 것이다. 이것을 식으로 표현하면 식(1)과 같다. 이와 같은 기존의 지형학적 특징 추출 방법에 이웃 화소의 평균값 차등화 방법을 추가하여 차등 값 $a_1, a_2, a_3, \dots, a_n$ 을 식(1)에 반영하면 식(2)와 같이 된다. $a_1, a_2, a_3, \dots, a_n$ 을 일일이 결정한다는 것은 매우 어려운 문제이므로 모든 변수에 대하여 식(3)과 같이 임의의 값 a_{th} ($=a_1, a_2, a_3, \dots, a_n$)를 동일하게 설정하고 a_{th} 을 일정한 크기로 증가시키면서 문자의 색상에 대한 특성을 반영한다. β -임계값은 문자의 서체와 관련이 있다고 했다. 즉, 문자의 획 폭과 관련이 있다. 따라서 그림3(a)의 마스크 크기를 증가시키면 문자의 획 폭에 대한 문제를 해결 할 수 있다. 그러면 마스크의 크기가 계속 커진다면 문제는 없는가? 여기에 대한 문제점은 문제의 복잡성이 증가하고 마스크 연산에 대한 처리 시간이 너무 많이 소요되는 문제점이 발생한다. 긴 처리 시간 문제는 그림 3(a)와 같이 표시된 부분만을 계산에 참여시킴으로써 처리 시간을 단축한다. γ -임계값은 문자의 크기와도 관련이 있다고 볼 수 있으나 문자 획과 문자 획간의 거리를 말하는 것으로 γ -임계값을 결정할 때 모든 문자에 대하여 처리가 가능하도록 보통 넓은 범위의 γ -임계값으로 정한다.



(a) 마스크 표현 형태



(b) 문자의 지형학적 표현 형태

그림3. 문자의 특징 마스크와 α , β 값과의 관계
Fig.3. relating α , β value and feature mask of character

$$\text{Feature-Point} = \begin{cases} 1 : \text{if}(A_1 > A_2 > A_3, \dots, A_n) \\ 0 : \text{otherwise} \end{cases} \dots \text{식(1)}$$

$$\text{Feature-Point} = \begin{cases} 1 : \text{if}((A_1 > A_2 + \alpha_1) \text{ and } (A_2 > A_3 + \alpha_2) \text{ and } (A_3 > A_4 + \alpha_3), \dots, (A_n > A_{n+1} + \alpha_n)) \\ 0 : \text{otherwise} \end{cases} \dots \text{식(2)}$$

$$\text{Feature-Point} = \begin{cases} 1 : \text{if}((A_1 > A_2 + a_{th}) \text{ and } (A_2 > A_3 + a_{th}) \text{ and } (A_3 > A_4 + a_{th}), \dots, (A_n > A_{n+1} + a_{th})) \\ 0 : \text{otherwise} \end{cases} \dots \text{식(3)}$$

입력 영상에 대해 식(2)에 의하여 구하여진 지형학적 특징점 추출 결과는 그림 4와 같다. 여기 문자 영역 추출 시간을 빠르게 하기 위하여 전체 영상에 대하여 특징점 추출을 하는 것이 아니라 일정 간격으로 특징점을 추출한다. 문자 영역은 모두 포함되어있고 문자 영역이 아닌 부분도 추출됨을 알 수 있다.

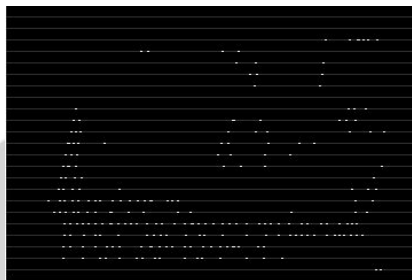


그림 4. 추출된 지형학적 특징점
Fig.4. extracted topographical features

III. MST(Minimu Spanning Tree)를 이용한 문자 영역 추출

그래프 $G=(V,E)$ 로 표현하며 정점(V)과 간선(E)으로 구성되어 있다. 2장에서 추출된 특징점을 정점으로 정의한다. 그리고 이들 정점과 연결될 수 있는 모든 간선들을 정의한다. 간선들이 연결될 수 있는 조건은 3가지(정점 간의 거리 임계치[Th-d], 정점간의 각 임계치[Th-a], 컬러의 차 임계치[Th-cd])를 만족하여야 연결될 수 있다. MST는 형성은 2개의 정점을 연결함으로써 형성시키는데 두 정점 사이에는 3가지 조건은 만족하면 정점과 정점을 연결하여 MST를 형성시켜 나간다. MST 형성 과정은 그림 5에서 알고리즘화하여 보여주고 있다. 그러면 3개의 임계 값을 어떻게 결정하는

가가 중요한 문제가 될 수 있다. 추출하고자 하는 문자 영역에 대한 비디오 영상의 특징을 알고 있다면 이러한 사전 지식을 이용하여 임계치를 정하면 된다. 만약 추출하고자 하는 문자의 특징을 알고 있지 않다면 기본값을 이용하여 결정한다. Th-d의 임계치는 그림 4의 특징점들 사이의 평균 거리로 결정한다. Th-a의 임계치는 0~45' 로 정한다. 이 의미는 수평적 문자 영역을 추출한다는 의미이다. Th-cd의 임계치 결정은 매우 어려운 문제로 볼 수 있다. 왜냐하면 이것은 같은 색상의 문자에서 얼마만큼의 범위를 같은 색으로 인정할 것인가와 관련이 있기 때문이다. 실험결과 Th-cd의 임계치는 50정도가 적정함을 알 수 있었다.

```

=====
Th-cd = default (unknown) / user define (known)
Th-a = default (unknown) / user define (known)
Th-d = default (unknown) / user define (known)

Gi = {V1} where i = 1 // number of graph

While (V is not empty)
{
v1 = one among V;
while (all Gi)
{
v2 = one element among Gi
ECD = Calculate Color Distance for E(v1 ,v2)

EA = Calculate Angle for E(v1 ,v2)
ED = Calculate Distance for E(v1 ,v2)

If(ECD < Th-cd && EA < Th-a &&
ED < Th-d && v1 is no-cycle in Gi)
Insert v1 into Gi
}
if( v1 is not included in any G i)
{
i = i + 1; // Creat New G i+1
Gi = v1;
}

Delete v1 from V
}
=====

```

where :: Th-cd : Threshold value for color distance
 Th-a : Threshold value for angle
 Th-d : Threshold value for distance

그림 5. MST를 이용한 문자 후보 영역 추출 알고리즘
 Fig.5. text extraction algorithm usin MST

그림 6은 특징점들에 대하여 MST를 연결한 것을 보여주고 있다. 이렇게 추출된 후보 영역은 연결된 정점의 개수 2개 이하거나, 영역의 크기가 문자 영역으로 보기에는 너무 작은 영역 등을 문자 후보 영역에서 제외시킨다. 그림 7에서 점선 원형에 있는 영역은 부적합한 정점 영역이며 이 영역을 제거한 후 추출된 문자의 후보 영역은 그림 7에서 보여주고 있다. 그림 8을 잡영 제거 후 MST에 의하여 형성된 문자 후보 영역을 보여주고 있다. 그림 8의 문자 후보 영역에서 문자 영역을 추출하기 위하여 이진화를 수행한다.

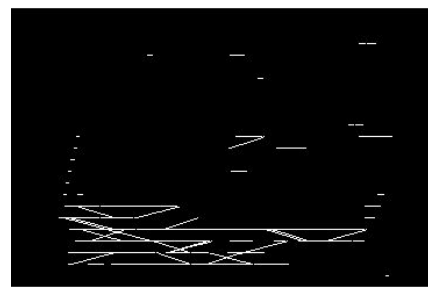


그림 6. MST에 의하여 형성된 문자 후보 영역
 Fig.6. Constructing candidate area by MST



그림 7. 후보 영역 잡영 제거
 Fig.7. deleting noises in candidate area

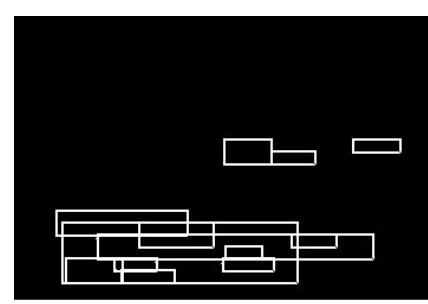


그림 8. 잡영 제거 후 추출된 문자 후보 영역
 Fig 8. extracted candidate area after deleting noise

그림 9는 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 안에있는 문자의 수가 3자 이하인 것은 문자 영역으로 간주하지 않는다. 그림 10은 검증후 추출된 문자 영역을 보여주고 있다.

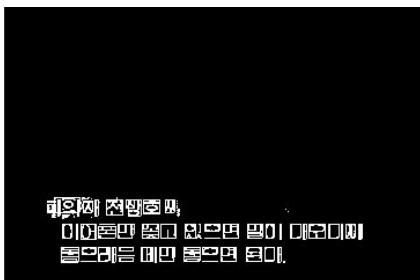


그림 9. 후보 문자 영역의 검증
Fig.9. verifying candidate areas of text

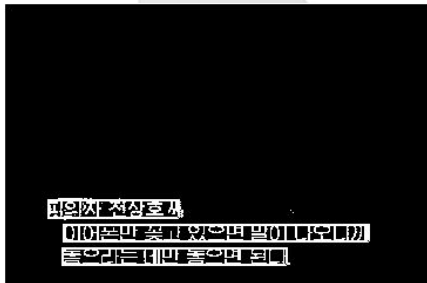


그림 10. 검증 후 추출된 문자 영역
Fig.10. extracted text area after verifying candidates

IV. 실험 결과 및 고찰

사용된 프로그램은 C++6.0이며 컴퓨터는 펜티엄 PC 3.0GHz CPU를 사용했다. 실험을 위하여 MPEG1 영상을 사용했으며 TV 영상을 영상 캡처보드를 사용하여 MPEG 영상을 획득했다.

실험 영상은 영화, 뉴스, 드라마 등을 사용했다. 실험 결과 표 1과 같이 뉴스 3972 프레임에 대하여 후보 영역 추출율은 100%, 문자 영역 추출율은 99%임을 알 수 있다. 영화의 경우 3679 프레임에 대하여 문자 영역 추출율은 99%, 문자 영역 추출율은 97%임을 알 수 있다. 드라마의 경우 3173 프레임에 대하여 후보 영역 추출율은 98%, 문자 영역 추출율은 96%임을 알 수 있었다. 실험 결과 MST를 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	3972	100	99
영화	3679	99	97
드라마	3173	98	96
스포츠	2934	99	98

IV. 결론

본 논문에서는 MST를 이용한 비디오 영상에서 문자 영역 추출하는 새로운 방법을 제안하였다. 비록 많은 실험을 수행하지 못하였지만 시스템 성능을 입증할 수 있었다. 향후 연구 방향은 컬러의 임계값(Th-cd)을 자동적으로 구할 수 있는 방법을 연구하여야 하며, 다른 군집화 방법을 이용한 문자 후보 영역 군집 방법에 대하여 연구하여 한다.

참고문헌

- [1] Shunji Mori, Ching Y. Suen, Kazuhiko Yamamoto, "Historical review of OCR research and Development", Proceeding of the IEEE, Vol.80, No.7, pp.1029-1058, July 1992.
- [2] K.Matsuo, K.Ueda, and M.Umeda, "Extaction of character string region on signboard from scene image using adaptive threshold method," IEICE, Vol.J80-D, No.6, pp.1617-1626, 1997.
- [3] Rainer Lienhart and Frank Stuber, "Automatic text recognition in videos," SPIE Storage and

Retrieval for Image and Video DB V, Vol. 3022, pp.368-378, Feb. 1997

- [4] Shoji Kurakake, Hidetaka Kuwano, Kazumi Odaka, "Recognition and visual matching of text region in video for conceptual indexing," SPIE Image and VideoProcessing IV, Vol. 2666, pp.180-188, Feb. 1996
- [5] J. Zhou, D.Lopresti, and T.Tasdizen, "Finding Text in color," Proc. of SPIE on Document Recognition V, pp.130-140, 1998.
- [6] B.L. Yeo and B. Liu, "Visual content highlighting via automatic extraction of embedded captions on MPEG compressed video", In Proc. SPIE Digital Video Compression : Algorithm and Technologies, Vol.2668, Feb. 1996.
- [7] Michael A. Smith and Takeo Kanade, "Video Skimming for Quick Browsing Based on Audio and Image Characterization,"Carnegie Mellon University,Technical Reprot CMU-CS-95-186, July 1995
- [8] Jun Ohya, Akio Shio and Shigeru Akamatsu, "Recogniton characters in scene images," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 16, No.2, pp.214-220, 1994
- [9] Byung Tae Chun, Younglae Bae and Tai-Yun Kim, "Text extraction in videos using Topographical features of character," The 8th IEEE Inter. Conf. on Fuzzy System (FUZZ-IEEE'99), Vol.2, pp.1126-1130, Aug., 1999.

저자 소개



전 병 태

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

관심 분야 : 멀티미디어 영상처리,
 영상처리, 동영상 처리, 컴퓨터 비
 전 등



김 영 인

1986년 한남대학교 전자계산학과
 (학사).
 1988년 명지대학교 전자계산학과
 (석사).
 1996년 명지대학교 컴퓨터공학과
 (박사).
 1990년~1991년 (주)한국컴퓨터
 기술연구소 연구원
 1996년~2006년 밀양대학교 컴퓨
 터공학부 부교수.
 2006년~현재 부산대학교 바이오
 시스템공학부 부교수.

관심 분야 : 바이오지능정보시스템,
 멀티미디어정보검색, 멀티미디어영
 상처리 등