

# 자막 자동 추출을 위한 강건한 자막 분리 알고리즘

## Robust text segmentation algorithm

### for automatic text extraction

정제희<sup>1</sup>, 정종면<sup>2</sup>

<sup>1</sup> 경기도 수원시 성균관대학교 공과대학

E-mail: gullingi@hotmail.com

<sup>2</sup> 전라남도 목포시 목포해양대학교 해양전자통신공학부

E-mail: jmjeong@mmu.ac.kr

#### 요 약

본 논문에서는 비디오에서 자막을 자동 추출하기 위한 강건한 자막 분리 알고리즘을 제안한다. 주어진 비디오에서 자막이 존재할 가능성이 있는 프레임에 대해 자막 후보 영역의 위치를 찾은 다음, 자막 후보 영역으로부터 강건하게 자막을 추출한다. 추출된 자막 후보 영역에 대해 Dampoint labeling을 수행하여 자막과 비슷한 색상을 갖는 배경을 제거하고, 마지막으로 기하학적 검증들 통해 최종적으로 자막 여부를 판별한다. 제안된 방법을 여러 장르의 비디오에 대해 적용한 결과 복잡한 배경을 갖는 비디오에서 자막을 강건하게 추출함을 실험을 통해 확인하였다.

**Key Words** : Text Extraction, text detection, text localization, text segmentation, geometrical verification, dampoint labeling, color polarity

## 1. 서론

디지털 비디오에서 광범위하게 사용되는 자막은 비디오 내용에 대한 시청각 정보를 보강하고, 다양한 부가 데이터를 비디오 이용자에게 제공할 수 있다. 따라서 비디오에 존재하는 자막을 자동 추출할 수 있다면, 비디오에 대한 내용기반 검색을 비롯한 여러 분야에서 유용하게 사용될 수 있다. 10여년 전부터 비디오로부터 자막을 자동추출하기 위한 연구가 광범위하게 이루어지고 있는데, 색상(color), 형태(shape), 질감(texture), 움직임(motion)등의 정보에 기반 하여 자막 후보 영역을 추출하고, 자막의 기하학적인 특징과, 인접한 프레임 사이의 일관성(consistency)을 이용하여 자막에 대한 검증을 수행한다.[1]-[4]

자막을 추출하기 위해서 대부분의 기존 방법들은 주어진 프레임에 자막이 존재하는지 여부를 판단하기 위한 자막 탐지(text detection) 단계와 자막이 탐지된 프레임에서 자막의 위치를 결정하는 단계(text localization). 찾은 자막 위치에서 자막과 배경을 분리하여 자막을 추출하는 단계(text segmentation)로 이루어져 있다[1].

[5]에서 제안된 방법은 자막의 획 정보를 이용하여 자막을 탐지하고, 모폴로지 연산을 이용하여 강건하게 자막을 추출하였다. 그러나 자막 분리 단계가 단순한 이진화에 의해 수행됨으로써 복잡한 배경을 갖는 비디오일 경우

오동작하는 경우가 있다.

한편, Lyu 등은 다국어가 포함된 비디오에서 자막을 추출하기 위한 알고리즘을 제안하였는데, Dampoint labeling을 사용하여 자막과 유사한 배경의 색상을 가진 배경을 제거하였으며 [3], Jiqiang 등은 자막의 최외각 테두리를 이루는 에지를 이용하여 자막의 색상에 무관하게 자막을 추출하는 방법을 제안하였다[4].

본 논문에서는 [5]의 알고리즘을 개선하여 자막 자동 추출을 위한 강건한 자막 분리 알고리즘을 제안한다. 논문의 이후 구성은 다음과 같다. 2장에서는 자막 탐지, 3장에서는 자막 위치 찾기에 대해 기술한 다음, 4장에서 자막 분리 방법을 설명한다. 5장에서는 찾아진 자막을 유효한 자막인가를 검증하는 기하학적 검증을 설명하고, 6장에서 실험 결과를 보인 후, 7장에서 결론을 맺는다.

## 2. 자막 탐지

[5]에서는 모폴로지 연산을 이용한 강건한 자막 추출 알고리즘을 제안하였는데, 본 논문에서는 [5]에서 제안된 알고리즘을 이용하여 자막 탐지 및 자막 위치 찾기를 수행한다. 먼저 자막 탐지 단계는 영상 안에 자막의 존재 여부를 판단하는 단계이다. 이를 위하여 주어진 입력 영상을  $M \times N$  크기의 블록단위로 분할한 다음 각 블록에 대하여, 획이 될 가능성이 높은 점, 즉 씨앗점을 탐색한다.

씨앗점 탐색은 획 경계부분의 에지방향을 이용하여 진행된다. 그림 1에서 보이는 바와 같이 각 획의 경계부분은 서로 대응되는 에지의 쌍이 일정 거리 이내에 존재하는데, 이런 특성



그림 1 획 경계 부분의 에지 방향을 만족하는 대응쌍을 씨앗점으로 고려한다.

한편, 일반적으로 디지털 영상에서 자막의 획을 구성하는 에지는 두 픽셀 이상의 폭을 갖는다. 그림 2에서 보이는 바와 같이 자막을 이루는 획의 경계는 대응되는 에지쌍이 여러 개 존재 할 수도 있다. 따라서 인접한 대응쌍들의 거리를 계산하여 최대 빈도수를 갖는 대응쌍들을 선택하고, 최대 빈도수를 갖는 대응쌍들 중 에지의 강도가 가장 강한 대응쌍을 씨앗점으로 결정한다.

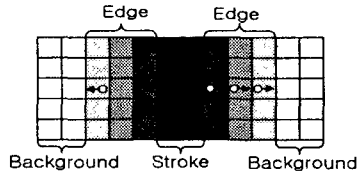


그림 2 복수개의 대응 쌍

이런 씨앗점들은 자막 영역에 집중적으로 나타나기 때문에, 주어진 블록 안에 임계치 이상의 씨앗점을 포함하지 않은 블록은 자막을 포함할 가능성이 낮다. 따라서 임계치 이상의 씨앗점을 갖지 못한 블록의 씨앗점은 모두 제거한다. 그런 다음 분할되었던 블록들을 통합하여 하나의 영상으로 만든 다음, 씨앗점들에 대해 모폴로지 폐쇄(closing) 연산과 모폴로지 개방(openning) 연산을 순차적으로 수행하여 인접한 씨앗점들은 통합하고, 고립된 씨앗점들을 제거하면 결과적으로 남아있는 씨앗점들은 자막의 획을 이루는 씨앗점이라 할 수 있다. 따라서 이 과정을 거친 후에도 씨앗점 영역이 남아 있는 프레임에는 자막이 존재할 가능성이 있다고 할 수 있다.

### 3. 자막 위치 찾기

자막 탐지 과정에서 나온 결과물인 씨앗점들은 자막뿐만 아니라 자막과 유사한 특성을 가진 씨앗점들을 포함하고 있다. 자막 위치 찾기(text localization)는 영상으로부터 자막과 유사한 영역에서 찾아진 씨앗점을 제거한 후 자막의 위치를 찾아내는 과정이다. 이를 위해 모폴로지 개방연산을 통하여 잡음들을 제거한 후, 잡음이 제거된 에지영역에 4-CC Labeling을 수행하여 자막 후보 영역에 존재하는 에지 영역들을 서로 분리한 다음, 분리된 영역 중에 씨앗점을 포함하지 않는 부분은 제거한다.

자막 위치 찾기는 위 단계에서 제거되지 않는 영역에 대한 수평 수직 프로젝션을 통해 수행한다. 프로젝션을 효율적으로 하기 위해선 각 분리된 자막지역의 프로젝션 정보는 피크를 이루는 것이 프로젝션에 유리하다. 따라서 모폴로지 폐쇄 연산을 다시 한번 수행하여 자막의 획 경계선 내부 영역을 채운 다음 프로젝션을 수행한다. 프로젝션은 반복적으로 이루어지며, 더 이상 프로젝션 영역의 정보 변화가 없을 때까지 이루어지게 된다.

프로젝션이 완료되면 자막을 포함할 확률이 높은 자막 후보 영역이 결정되며, 이 영역은 바운딩 박스 형태를 갖는다.

### 4. 자막 분리

자막 분리(text segmentation) 단계는 바운딩박스에 있는 자막과 배경을 서로 분리하는 단계이다. 이를 위하여 [5]에서는 바운딩 박스의 픽셀에 대한 이진화를 수행하였다. 그러나 자막의 색상은 주어지지 않기 때문에 이진화된 영상에서 어느 쪽을 자막 영역으로 선택하느냐에 대한 고려가 필요하다. 또한 배경이 복잡한 경우 자막과 배경 영역이 명확하게 구분되지 않는 경우가 종종 발생한다. 이를 고려하여, 본 논문에서는 Dampoint labeling[3]과 Color polarity 기법[4]을 이용한다.

이진화를 위하여 자막 위치 찾기 단계에서 얻어진 바운딩박스의 가로 세로 방향으로 약간 확장된 영역을 고려한다. 자막 탐지 단계에서 추출된 씨앗점은 획의 경계선에 존재하기 때문에 확장된 바운딩박스에 존재하는 씨앗점 밝기의 평균값은 바운딩박스 영역을 이진화하기 위한 임계값으로 사용할 수 있다.

이진화 작업이 끝난 영상에서 흰색 또는 검정색 영역을 자막 영역으로 결정해야 하는데, Color polarity 기법을 사용하였다[4]. 이 방법은 그림 3의 3X3의 Edge Mask  $K_b$ 와  $K_w$ 를 각 픽셀  $P(x,y)$ 에 적용하여, 식 1의 조건에 따라 White, None, Black Edge로 구분한다. 이렇게 만들어진 에지로 이루어진 에지영상과 자막을 이루는 최외각 에지의 정보를 제거한 에지 영상의 White, Black Edge의 픽셀의 비율을 계산하고, 실험적 자료를 이용하여 자막의 색상이 밝은지 어두운지를 판단한다. 만약 자막이 어두운 색으로 판단되면 바운딩박스의 역상화를 진행한다.

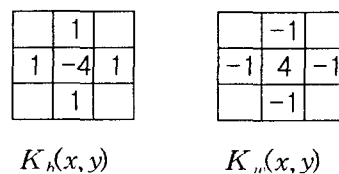


그림 3 Edge Mask

$$P(x,y) \begin{cases} \text{White Edge, } K_w(x,y) > 0 \\ \text{Black Edge, } K_b(x,y) > 0 \\ \text{Non Edge, } K_w(x,y) \leq 0 \text{ and } K_b(x,y) < 0 \end{cases} \quad \text{- 식 1}$$

이진화 영상에서 자막을 이루는 픽셀과 배경 픽셀의 밝기가 유사한 경우는 복잡한 배경에서 종종 등장하는데 이러한 경우엔 배경 픽셀까지 자막 픽셀에 포함되는 경우가 발생한다. 이런 문제를 해결하기 위해서 Dampoint labeling을 이용한다[3]. Dampoint labeling은 자막을 이루는 획의 두께는 일정하다는 특징을 이용하는데, 식2의 조건에 따라 획의 두께와 다른 크기를 갖는 영역을 제거한다.

$$\text{Damp Point} = (x,y) \{ B(x,y) = \text{White} \wedge \text{MIN} - W < \text{MIN}[H - \text{Len}(x,y), V - \text{Len}(x,y)] < \text{MAX} - W \} \quad \text{- 식 2}$$

여기에서 White란 이진화 영상의 픽셀의 색상으로 자막의 색상이 밝은 경우를 말한다. MIN-W와 MAX-W는 자막을 이루는 획의 최소 최대 간격을 말하며, H-Len과 V-Len은 각 픽셀의 상하 좌우로 연속되는 밝은 색 픽셀의 개수를 의미한다.

그림 4(a)는 자막은 밝은 색, 배경은 어두운 색으로 나타난 자막을 포함한 이진화된 영상이다. 그림 4(b)는 Dampoint labeling을 수행하지 않고, 바운딩박스의 상하좌우로 바운딩박스의 1/8의 크기만큼 확장된 바운딩 박스의 영역에서의 밝은 픽셀로부터 Flood-Fill 알고리즘을 사용한 결과 's'라는 자막이 삭제됨을 보여준다. 이러한 자막과 픽셀과 유사한 색상을 지닌 배경 픽셀이 자막과 인접하는 경우엔 Flood-Fill에 의해 글자가 사라질 수 있다. 그림 4(c)는 각 픽셀을 식 2의 조건으로 Dampoint labeling을 수행한 결과이다. 색상이 회색인 픽셀은 Dampoint labeling이 된 픽셀이다. 그림 4(d)에선 그림 4(c)에서 확장된 바운딩박스의 영역의 밝은 색의 각 픽셀로부터 Flood-Fill 알고리즘을 적용하고, 어두운 색을 가진 픽셀을 제외한 나머지 픽셀들을 밝은 색으로 변환한 영상이다. 본 논문에서는 [3]에서 제안한 Dampoint labeling의 MIN-W와 MAX-W를 씨앗점을 탐색하기 위한 대응쌍의 검색 범위로, 확장된 바운딩 박스의 영역은 주어진 비디오에서 나타날 수 있는 자막 높이의 1/8로 하였다. Dampoint labeling 단계를 거치면, 배경으로부터 분리된 자막 후보영역을 얻는다.

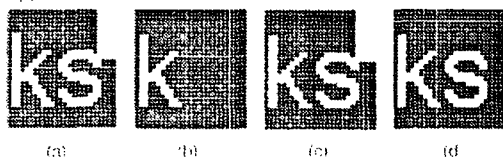


그림 4 Dampoint labeling 진행

### 5. 기하학적 검증

기하학적 검증 단계에서는 배경으로부터 분리된 자막 후보 영역 중 기하학적으로 자막이 될 수 없는 영역을 제거함으로써 자막을 최종적으로 추출한다.

자막 위치 찾기 단계에서 얻은 바운딩박스는 자막과 인접하게 존재하는 잘못된 배경 영역을 포함하는 경우가 있기 때문에, 자막 분리단계를 거친 자막 후보 영역에 대해 다시 프로젝션을 수행하여 최종적인 바운딩 박스를 생성한다. 이 바운딩 박스는 글자 영역과 글자 사이에 나타나는 배경만을 포함하기 때문에, 바운딩박스 내부의 글자대 배경 픽셀은 일정한 임계값 이상을 갖는다. 본 논문에서는, 실험적으로 임계 비율을 얻은 다음, 자막대 배경 영역의 비율이 임계치를 벗어날 경우, 자막이 아니라고 판단하고 바운딩박스의 자막정보를 제거한다.

한편 비디오에는 자막의 획과 유사한 구조를 갖는 인공 구조물이 나타나는 경우가 매우 많은데, 이런 경우의 바운딩박스는 매우 큰 영역을 갖는다. 따라서 바운딩박스의 높이가 매우 큰 경우는 인공 구조물에 의한 영역으로 간주하고 바운딩박스의 자막 정보를 제거한다. 수평 방향으로 길게 나타나는 바운딩박스는 비디오 자막의 정렬 방향을 고려했을 때 인공 구조물이라고 단정하기 힘들다. 또한, 자막을 둘러싸고 있는 직선형의 테두리의 경우는 자막과 유사한 특성인 대응쌍과 인접한 씨앗점에 의해 추출 되기 때문에, 매우 작은 높이의 바운딩박스를 갖는다. 이러한 영역을 제거하기 위해 너무 작은 높이의 바운딩박스도 제거한다.

마지막으로 바운딩박스의 가로 세로의 길이를 구한 다음, 바운딩박스의 가로세로 비율(aspect ratio)을 이용하여 잘못된 바운딩박스를 제거한다.

### 6. 실험 결과

본 논문에서 제안한 알고리즘을 타당성을 보이기 위해 뉴스, 뮤직 비디오, 스포츠 등의 방송용 영상에 대한 실험을 수행하였다.

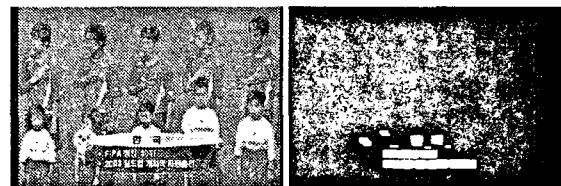


그림 5 . 입력 영상

그림 6. 모폴로지 폐쇄 연산 결과

