

디지털 비디오를 위한 획기반 자막 추출 알고리즘

A Stroke-Based Text Extraction Algorithm for Digital Videos

정종면* · 차지훈** · 김규현***

Jong-Myeon Jeong, Jihun Cha and Kyuheon Kim

* 목포해양대학교 해양전자통신공학부

** 한국전자통신연구원 전파방송연구단

*** 경희대학교 전자정보대학

요 약

본 논문에서는 디지털 비디오를 위한 획기반 자막 추출 알고리즘을 제안한다. 제안된 알고리즘은 자막 탐지, 자막 위치 찾기, 자막 분리 단계와 분리된 자막에 대한 기하학적 검증 과정으로 구성된다. 자막 탐지 단계는 연속적으로 입력되는 프레임 중 자막이 존재하는 프레임을 찾는 단계로써, 주어진 프레임으로부터 자막이 될 가능성이 높은 점, 즉 씨앗점을 추출한 다음 씨앗점에 대하여 모폴로지 연산을 수행한다. 자막 위치 찾기 단계는 자막이 존재하는 프레임에서 자막의 위치를 찾는 단계로써, 씨앗점을 포함하는 에지에 대한 모폴로지 연산과 프로젝션을 통해 수행된다. 자막 분리 단계에서는 자막과 배경의 색상 분포와 복잡한 배경을 고려하여 자막을 강건하게 분리한다. 마지막으로 자막에 대한 사전 정보를 이용하여 분리된 자막에 대한 기하학적 검증 과정을 수행하여 최종 결과를 얻는다.

키워드 : 자막 추출, 자막 탐지, 자막 위치 찾기, 자막 분리, 씨앗점

Abstract

In this paper, the stroke-based text extraction algorithm for digital video is proposed. The proposed algorithm consists of four stages such as text detection, text localization, text segmentation and geometric verification. The text detection stage ascertains that a given frame in a video sequence contains text. This procedure is accomplished by morphological operations for the pixels with higher possibility of being stroke-based text, which is called as seed points. For the text localization stage, morphological operations for the edges including seed points are adopted followed by horizontal and vertical projections. Text segmentation stage is to classify projected areas into text and background regions according to their intensity distribution. Finally, in the geometric verification stage, the segmented area are verified by using prior knowledge of video text characteristics.

Key Words : Text extraction, Text detection, Text localization, Text segmentation, Seed point

1. 서 론

자막은 비디오 데이터의 시청각적인 정보를 보강하고, 부가 정보들을 사용자에게 제공하며, 방대한 멀티미디어 데이터의 내용을 예측할 수 있는 중요한 단서를 제공한다. 또한 색상(color), 형태(shape), 질감(texture), 움직임(motion) 등의 특징을 이용하여 내용기반 비디오 검색 시스템을 구축하는 것에 비하여 자막 정보를 이용하면 그 비용과 정확성이 더 우수하고 사용자가 원하는 데이터를 찾는 데 보다 용이하다. 따라서 주어진 비디오로부터 자막 정보를 추출하기 위한 연구가 10여년 전부터 꾸준히 이루어지고 있다[1]-[6].

자막을 추출하기 위해서는 일반적으로 동영상에 자막이 존재하는지 여부를 판단하는 자막 프레임 추출(text de-

tection) 단계, 자막 프레임이 존재하는 프레임들로부터 자막의 최소 영역(minimum text region) 위치를 찾아내는 단계(text localization), 그리고 추출된 자막 최소 영역에서 자막을 분리(text segmentation)하는 단계를 거친다[1].

자막 탐지 단계에서는 자막을 일종의 질감(texture)로 간주한 다음 웨이블릿 변환(wavelet transform)[7], 이산 여현 변환(Discrete Cosine Transform)[8], 가보 필터(Gabor filter)[9], 공간적 변위(spatial variance)[10] 등의 특징을 추출하고, 이를 신경망(neural network)[11]이나 SVM(Support Vector Machine)[6]과 같은 패턴 분류기를 이용하여 자막과 유사한 질감을 갖는 영역을 탐지하거나, 자막을 일정한 색상을 갖는다고 가정하고 일정한 색상을 갖는 영역을 탐지한 후, 영역의 연결성을 검사하는 방법이 있다[6][11].

자막 위치 찾기 단계는 주어진 작은 자막 영역들을 통합하여 하나로 병합해 나가는 상향식 접근 방법[10]과, 전체 영역을 분할하여 작은 영역으로 나누어 자막 위치를 찾는 하향식 접근 방법이 있다[12][13]. 자막 분리 단계에서는 자막 영역을 이루는 픽셀은 다른 영역과는 다른 색상을 일관되게 갖는다는 가정하에 배경과 자막을 분리해 낸다[14].

접수일자 : 2007년 4월 1일

완료일자 : 2007년 6월 4일

감사의 글 : 본 연구는 정보통신부 및 정보통신연구진흥원의 IT신성장동력핵심기술개발사업의 일환으로 수행하였음.

지난 10여년 동안 자막 추출을 위한 많은 연구 성과가 있었으나 다양한 장르의 비디오에 적용하기에는 여전히 많은 어려움이 존재한다. 즉, 질감에 기반한 방법은 복잡한 배경이 존재하는 경우 오동작하기 쉬우며, 방대한 폰트, 다양한 배경, 그리고 자막에 대한 특수효과 등으로 인해 신경망이나 SVM을 적용하기 위한 training이 어렵다. 또한 색상이나 밝기 정보를 이용한 방법은 대부분의 범용 비디오에서 자막의 색상이 다양하게 나타나기 때문에 자막 영역에 대한 색상 정보를 추출하는 것이 선행되어야 한다.

본 논문에서는 복잡한 배경을 갖는 비디오로부터 자막의 색상, 위치 등에 무관하게 자막을 강건하게 추출하기 위한 알고리즘을 제안한다. 본 논문의 이후 구성은 다음과 같다. 먼저 2장과 3장에서는 획에 기반한 자막 탐지와 자막 위치 찾기 방법을 제안한다. 4장에서는 밝기 정보를 이용한 자막 분리 방법을 5장에서는 자막의 기하학적인 검증에 대해 설명한 다음 6장에서 실험결과를 보이고 7장에서 결론을 맺는다.

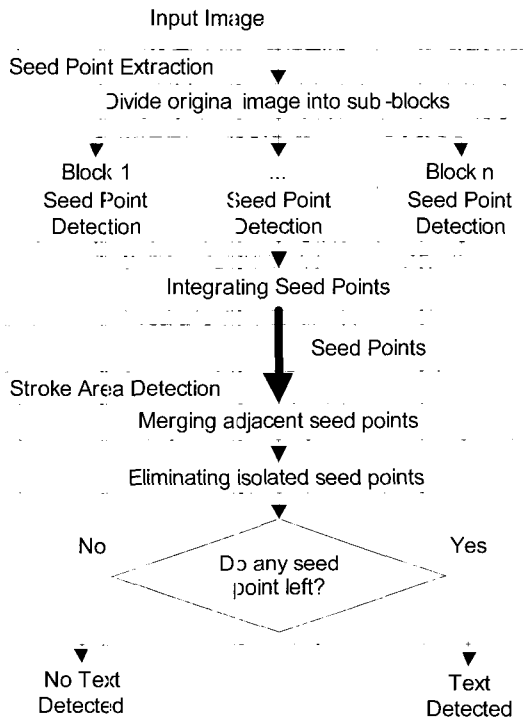


그림 1. 자막 탐지 단계의 블록도.
Fig. 1. Block diagram of text detection process.

2. 자막 탐지

비디오에 존재하는 자막은 언어, 폰트, 경계선 그리고 그림자 등 자막에 대한 특수 효과 등으로 인해 다양한 형태를 갖는다. 비디오에 존재하는 자막 정보는 경우에 따라서 단어 단위로 붙어서 나타나기도 하고 한 개의 문자만 떨어져서 나타나는 경우도 있다. 따라서 다른 영역과 자막 영역을 구별하기 위한 변별력 있는 특징(feature)을 결정하는 일은 어려운 일이다.

본 논문에서 제안하는 자막 추출 알고리즘은 획에 기반한 자막 탐지(text detection)와 자막 위치 찾기(text localization), 밝기 정보를 이용한 자막 분리(text segmentation),

그리고 자막 검증 등으로 이루어진다. 자막 탐지는 그림 1에서 보이는 바와 같이 씨앗점 추출과 획 후보 영역 탐색 과정으로 구성된다.

씨앗점이란 획이 될 가능성이 높은 점을 의미하는데, 씨앗점이 많이 나타나는 지역은 자막이 될 가능성이 높다고 할 수 있다. 한편, 자막은 한 개 이상의 획이 모여 문자를 이루고, 한 개 이상의 문자가 모여서 단어를 이룬다. 따라서 본 논문에서는 먼저 획이 될 가능성이 높은 점을 씨앗점(seed point)으로 추출한 다음 모폴로지 연산을 이용하여 서로 인접한 씨앗점들은 통합하고, 고립되어 나타나는 씨앗점은 제거함으로써 자막 탐지과정을 수행하였다

2.1. 씨앗점 탐색

자막은 자막을 구성하는 획의 경계선을 이루는 부분에서 서로 쌍을 이루고, 이 쌍은 항상 반대 방향 성분을 갖는다. 그림 2는 한글과 영문자의 획과 획의 경계선을 이루는 에지를 보이고 있는데, 그림에서 보이는 바와 같이 획을 이루는 임의의 에지는 일정한 거리 이내에 자신의 방향과 정반대 방향의 에지를 대응쌍으로 갖고, 서로 대응되는 에지 사이의 거리는 획의 폭 크기와 같다. 자막 영역이 아닌 영역에서 나타나는 에지에서 이런 대응관계가 나타날 가능성은 희박하며, 자막 영역에서는 이런 대응관계가 집중적으로 나타난다.

본 논문에서는 이런 대응관계가 국부적으로 집중적으로 나타나는 영역을 찾기 위하여 주어진 영상에서 에지를 추출한 다음, 에지 영상을 $N \times M$ 크기의 블록으로 분할한다. 그런 다음 블록에 존재하는 에지에서 일정 거리 이내에 존재하면서 반대 방향 성분을 갖는 에지를 주어진 에지에 대한 대응쌍으로 선택한다. 대응쌍이 많이 나타나는 블록은 자막이 존재할 가능성이 높은 블록이라고 할 수 있다.



그림 2. 획 경계부분의 에지 방향
Fig. 2. Directions of edges at stroke boundaries.

한편 디지털 영상의 특성상 에지의 두께는 두 픽셀 이상인 것이 일반적이다. 따라서, 에지의 방향만을 이용하여 획의 대응쌍을 찾으면 한 에지에 복수개의 대응쌍이 존재할 수 있으며, 이 경우에는 복수개의 대응쌍 중 가장 최적의 대응쌍을 선택하는 과정이 필요하게 된다. 그림 3은 2픽셀 두께의 획 영역을 중심으로 3픽셀 두께를 갖는 좌, 우의 에지를 보이고 있다. 그림 3에서 보이는 바와 같이 왼쪽 중앙의 에지 방향은 왼쪽을 향하고 있는데, 이에 대응하는 에지가 획 영역의 오른쪽에 있는 3개의 에지 중 어떤 것인지 결정하는 것이 어렵다. 반대 방향을 갖는 에지 중 가장 가까운 곳의 에지를 대응쌍으로 선택하는 방법을 생각할 수 있으나, 대부분의 경우 자막 영역의 화질이 매우 열악하고, 경계선에서 에지가 단일 에지로 존재하지 않기 때문에 단순히 가까운 곳의 에지를 대응쌍으로 선택하는 방법은 신뢰하기 힘든 결과를 출력할 수 있다. 따라서 여러 개의 대응쌍 중 최적의 대응쌍을 선택하기 위한 적절한 기준이 필요하다. 본 논문에서는 대응쌍 간의 거리, 즉 획의 폭을 먼저 계산한 다음, 획의 폭과 다른 거리를 갖는 대응쌍은 제거한다.

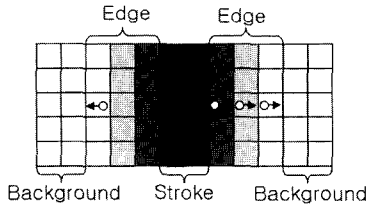


그림 3. 복수개의 대응쌍
Fig. 3. Multiple numbers of possible corresponding points.

획의 폭을 계산하기 위하여 주어진 에지들에 대한 대응쌍들을 모두 구한 다음 그 대응쌍의 거리에 대한 히스토그램을 생성한다. 잡음을 비롯한 여러 가지 요인에 의해 왜곡된 에지의 대응쌍의 거리들은 랜덤한 분포를 갖지만, 획을 이루는 에지들 중 왜곡되지 않은 대응쌍들은 일정한 거리를 유지하기 때문에, 대응쌍의 거리에 대한 히스토그램에서 최대 빈도(mode)를 갖는 거리는 획의 폭(width)에 해당하는 거리라고 할 수 있다.

획의 폭과 다른 거리를 갖는 대응쌍은 제거하고, 남아있는 대응쌍 중에서 에지의 강도를 고려하여 최적의 대응쌍을 선택한다. 즉, 에지 방향과 같은 방향으로 연속적으로 나타나는 동일한 거리의 대응쌍 중, 에지들의 강도의 합이 가장 강한 대응쌍을 최적의 대응쌍으로 선택한다. 그림 4는 씨앗점 탐색을 위한 흐름도를 보인다.

본 논문에서 제안하는 씨앗점 탐색 알고리즘을 요약하면 다음과 같다. 먼저 주어진 입력 영상을 일정한 크기를 갖는 블록으로 분할한다. 각 블록에 대하여 에지의 방향성을 이용하여 에지에 대한 최적의 대응쌍들을 찾고 그 대응쌍들이 블록 내에 임계치 이상 존재하는지를 판단한다. 만약 최적의 대응쌍이 임계치 이상 존재한다면 그 블록에 존재하는 대응쌍들은 획 후보 영역의 경계선을 이루는 씨앗점으로 남겨두고, 그렇지 않다면 그 블록에 존재하는 모든 대응쌍은 버린다. 남겨진 대응쌍들은 자막 후보 영역을 탐색하기 위한 씨앗점이 된다.

2.2. 획 후보 영역 탐색

자막은 한 개 이상의 획들이 조합되어 만들어진다. 따라서 획 경계 씨앗점들이 자막으로부터 얻어진 것이라면 자막 크기 정도의 영역에 집중적으로 나타나며, 자막과 유사한 모양의 영상에서 얻어진 것이 아니라면 고립되어 나타난다. 이런 특징을 이용하여 본 논문에서는 획 경계 씨앗점 탐색 단계에서 얻어진 씨앗점에 대한 모폴로지 연산을 통해 자막 후보 영역을 탐색한다. 먼저 주어진 씨앗점에 대해 모폴로지 폐쇄 연산(morphological closing)을 수행하면 고립되어 있는 씨앗점, 즉 잘못된 씨앗점들은 변화가 없는 반면, 실제 자막을 이루는 씨앗점들은 일정한 영역에 집중적으로 분포하기 때문에 서로 연결되어 한 개의 영역으로 통합되어 자막 후보 영역이 된다. 그런 다음, 다시 모폴로지 개방 연산(morphological opening) 연산을 수행하면 고립되어 있는 영역은 제거되고, 자막 영역에 포함되어 있던 획 영역은 변화 없이 그대로 존재한다[15]. 따라서 주어진 프레임에서 모폴로지 폐쇄 연산과 개방 연산을 수행한 후에도 남아있는 획 후보 영역이 존재한다면 자막이 탐지되었다고 할 수 있다.

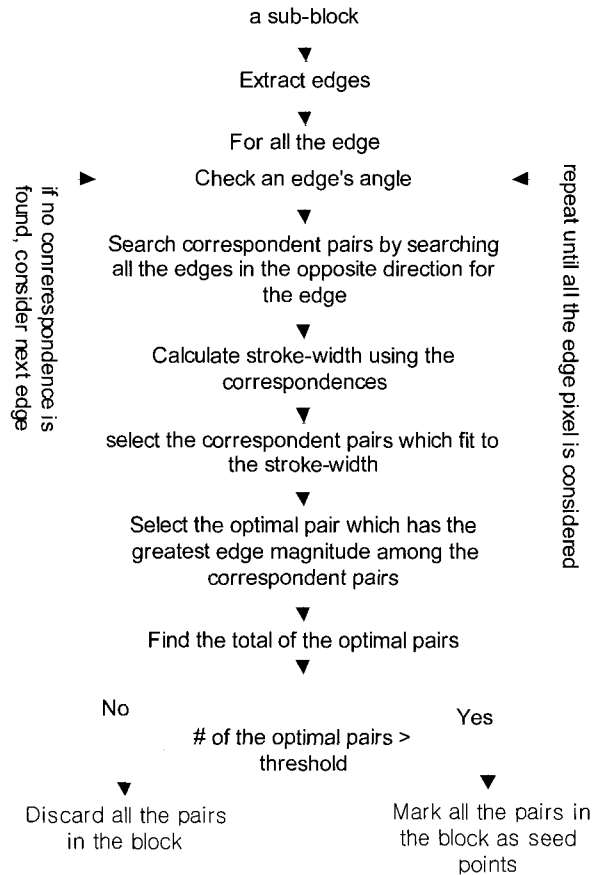


그림 4. 씨앗점 탐색의 흐름도
Fig. 4. Flowchart of seed point detection process.

3. 자막 위치 찾기

획 후보 영역이 존재하는 프레임에 대하여 자막 위치 찾기 과정을 수행한다. 자막 위치 찾기는 주어진 프레임에서 자막이 존재하는 영역을 정확하게 찾아내는 것을 목표로 한다.

획 후보 영역에는 복잡한 배경과 잡음으로 인해 획이 아닌 에지들이 존재하기 때문에 먼저 획 후보 영역에 존재하는 에지에 대해 모폴로지 개방 연산을 수행하여 잡음을 제거한다. 잡음이 제거된 에지 영역들에 대해 4-CC labeling을 수행하여 획 후보 영역에 존재하는 에지 영역들을 서로 분리한 다음, 서로 분리된 에지 영역들 중 씨앗점을 포함하는 영역을 제외하고 모두 제거한다. 이상적인 경우, 이 과정을 거친 에지 영역은 획의 경계선을 이루고 있다고 할 수 있으며, 경계선 영역의 내부는 획 영역이기 때문에 일정한 밝기 분포를 갖는 영역이 존재한다. 그러나 많은 경우 복잡한 배경과 잡음의 영향으로 획이 아닌 영역도 포함한다. 자막은 일정한 크기를 갖으며, 여러 개의 획으로 구성되기 때문에 획의 경계선을 이루고 있는 에지에 대해 수평, 수직 프로젝션을 수행하면, 자막의 특징을 갖지 않는 영역을 제거할 수 있다.

한편, 자막 영역을 효과적으로 찾기 위해서는 자막 영역에서 프로젝션의 peak가 강하게 나타나는 것이 바람직하다. 따라서 본 논문에서는 획 경계선의 내부를 채운 다음 프로젝션을 수행함으로써 프로젝션이 효과적으로 수행될 수 있도록

하였다. 이를 위하여 모폴로지 연산을 다시 한번 수행하여 인접한 획 경계선들을 통합한다. 획 경계 영역에 대해 모폴로지 폐쇄 연산을 수행하면 서로 인접한 획 경계는 자막 영역으로 병합되며, 획 경계 내부 영역도 채워지게 된다. 그러나 획 영역과 유사한 구조를 갖는 배경이 있는 경우에는 획 영역은 자막으로 병합되지 않고 유지된다. 그런 다음 수평, 수직 방향의 프로젝션을 반복적으로 수행하면, 잡음이나 배경에 의한 획 영역을 효과적으로 제거할 수 있으며, 최종적으로 자막 후보 영역만 남는다. 찾아진 자막 후보 영역은 바운딩박스(bounding box)의 형태로 자막 분리 단계로 전달된다.

4. 자막 분리

자막 위치 찾기 수행의 결과 얻어진 자막 후보 바운딩박스는 획의 경계선과 자막 사이의 배경 영역을 포함한다. 자막 분리 단계는 자막 후보 바운딩박스로부터 자막을 분리하여 추출하는 단계인데, 자막은 그 색상이 어둡거나 밝은색으로 단정할 수 없기 때문에 이진화된 영역 중 자막 영역을 선택해야 할 뿐 아니라, 자막 색상과 유사한 색을 포함하는 복잡한 배경을 갖는 경우에 대한 고려가 필요하다.

본 논문에서는 자막을 분리하기 위하여 먼저 이진화를 수행한 다음, 이진화된 영역에서 자막 색상을 결정하고 자막과 유사한 색을 갖는 배경을 제거함으로써 자막 후보 바운딩박스에서 자막을 분리해 낸다.

본 논문에서는 이진화를 위한 임계값을 구하기 위하여 자막 탐지 단계에서 얻은 씨앗점의 밝기값을 이용한다. 즉, 씨앗점은 획 경계선의 중앙에 위치하며 가장 강한 에지 강도를 갖기 때문에 주어진 자막 후보 바운딩박스에 존재하는 씨앗점 밝기값의 평균을 바운딩박스 영역에 대한 이진화의 임계값으로 사용한다.

자막 후보 바운딩박스에 대해 이진화를 수행하면 바운딩박스 영역은 자막과 배경의 두 영역으로 분할된다. 두 영역 중 어느 영역이 자막인지에 대한 판단은 Song 등이 제안한 color polarity 기법을 사용하였다.[16]

Color polarity 기법은 이진화된 자막 후보 영역에서 자막의 색상에 대한 사전 정보 없이 자막과 배경 영역을 구분하기 위한 기법으로써, 그림 5의 3X3의 에지 마스크 K_b 와 K_w 를 이진화된 픽셀 $P(x, y)$ 에 적용하여, 식 1의 조건에 따라 *White*, *None*, *Black Edge*로 구분한다.

이렇게 만들어진 에지영상과 자막을 이루는 최외각 에지의 정보를 제거한 에지 영상의 *White*, *Black Edge*의 픽셀의 비율을 계산하고, 이 비율에 대한 실험적 자료를 이용하여 자막의 색상이 밝은지 어두운지를 판단한다. 만약 자막이 어두운 색으로 판단되면 바운딩박스의 역상화를 진행하여 자막이 항상 밝은색이 되도록 한다.

	1	
1	-4	1
	1	

$K_b(x, y)$

	-1	
-1	4	-1
	-1	

$K_w(x, y)$

그림 5. Color polarity를 위한 에지 마스크.
Fig. 5. Edge Masks for color polarity.

$$P(x, y) \begin{cases} \text{White Edge, } K_w(x, y) > 0 \\ \text{Black Edge, } K_b(x, y) > 0 \\ \text{None Edge, } K_w(x, y) \leq 0 \text{ and } K_b(x, y) \leq 0 \end{cases} \quad (\text{식1})$$

한편 복잡한 배경을 갖는 자막 후보 영역에서는 자막과 배경 픽셀의 밝기가 유사한 경우가 있는데, 이런 경우에는 이진화된 자막영역에 배경 픽셀이 함께 포함되는 경우가 발생한다. 본 논문에서는 복잡한 배경으로부터 자막을 강건하게 분리하기 위하여 Lyu 등이 제안한 dampoint labeling 기법을 이용한다[2]. Dampoint labeling 기법은 자막 후보 바운딩박스의 외곽에서부터 flood fill 알고리즘을 이용하여 배경 영역을 제거하는데, 통상적인 flood-fill 알고리즘은 그림 6(a), 그림 6(b)에서 보이는 바와 같이 자막 영역도 배경으로 취급하여 제거한다. 따라서 Lyu 등은 자막을 이루는 획의 두께는 일정하다는 점을 이용하여 주어진 픽셀이 획의 일부분 인지를 판단하여 dampoint로 마크한다. 식 2는 주어진 픽셀이 dampoint 인지 여부를 판단하기 위한 조건을 보이고 있다.

$$\text{Damp Point} = (x, y) | B(x, y) = \text{White} \wedge \text{MIN-W} \leq \text{MIN}[H\text{-Len}(x, y), V\text{-Len}(x, y)] \leq \text{MAX-W} \quad (\text{식2})$$

여기에서 *White*란 고려 중인 이진 영상의 픽셀이 밝은 것을 의미하고, *MIN-W*과 *MAX-W*는 자막을 이루는 획의 최소, 최대 두께를 의미하며, *H-Len(x, y)*와 *V-Len(x, y)*는 (x, y) 좌표를 중심으로 상하 좌우로 연속되는 밝은 색 픽셀의 개수를 의미한다.

주어진 자막 후보 바운딩박스 영역을 약간 확장한 다음 최외곽의 밝은 픽셀부터 flood-fill 알고리즘을 적용하되, dampoint를 만나면 flood-fill을 멈춘다.

그림 6(c)는 주어진 자막 후보 영역에 대해 dampoint labeling을 수행한 결과로써, 회색으로 표시된 픽셀이 dampoint로 판정된 픽셀이다. 그림 6(d)는 그림 6(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-W*과 *MAX-W*를 씨앗점을 탐색하기 위한 대응쌍의 검색 범위로, 확장된 바운딩박스의 영역은 주어진 비디오에서 나타날 수 있는 자막 높이의 1/8로 하였다. Dampoint labeling 단계를 거치면, 배경으로부터 분리된 자막 영역을 얻는다.

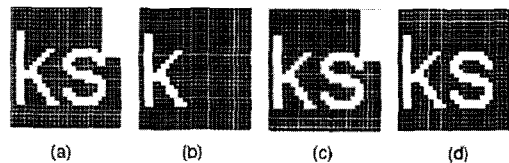


그림 6. Dampoint labeling 진행.
Fig. 6. Dampoint labeling process.

5. 기하학적 검증

비디오에는 자막의 획과 유사한 질감을 갖는 배경 영역들이 존재한다. 특히 인공 구조물이 나타나는 비디오에는 획과 유사한 구조가 나타나는 경우가 많이 있어, 자막 분리 단계의 dampoint labeling 기법으로 제거되지 않는다. 따라서 본

논문에서는 자막 분리 단계를 거친 자막 영역에 대해 기하학적 검증을 통해 기하학적으로 자막이 될 수 없는 영역을 제거하여 최종 결과를 얻는다.

자막 위치 찾기 단계에서 얻은 자막 후보 바운딩박스는 자막과 인접한 잘못된 배경 영역을 포함하는 경우가 있기 때문에 자막 분리단계를 거친 자막 영역에 대해 다시 프로젝션을 수행하여 자막 영역에 대한 바운딩박스를 생성한다. 이 바운딩박스는 자막 영역과 획 사이의 배경 영역만을 포함하기 때문에, 바운딩박스 내부의 글자대 배경 픽셀 비율은 0이나 1에 극단적으로 가깝지 않다. 따라서 본 논문에서는 바운딩박스의 글자대 배경 픽셀 비율이 임계치를 벗어나면 바운딩박스의 자막정보를 제거한다.

한편 바운딩박스의 높이(height)는 자막의 높이와 같은데, 이는 자막 획의 두께와 밀접한 관계가 있다. 따라서 본 논문에서는 자막 획의 두께를 고려하여 바운딩박스의 높이가 지나치게 크거나 작은 경우에는 정상적인 자막으로 볼 수 없기 때문에 바운딩박스의 자막 정보를 제거한다. 바운딩박스의 높이가 지나치게 큰 경우는 비디오에 질감정보가 풍부한 인공 구조물의 경우가 많으며, 너무 작은 경우는 자막의 가시성 향상을 위하여 자막에 삽입한 인공적 배경의 경계선인 경우가 많다.

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

6. 실험 결과

본 논문에서 제안하는 자막 추출 기법의 타당성을 입증하기 위하여 뉴스, 스포츠, 엔터테인먼트 등의 비디오에 대한 실험을 수행하였다. 실험에 사용한 영상은 720*480으로 해상도, 256단계의 농담값을 갖는다. 그림 7은 제안한 알고리즘을 실행한 각 단계별 결과이다.

그림 7(a)는 입력 영상을 보이고 있으며, 그림 7(b)는 입력 영상의 에지를 추출한 모습이다. 그림 7(c)의 씨앗점은 모폴로지 폐쇄 연산을 통해 그림 7(d)와 같이 통합된 다음, 모폴로지 개방 연산 및 모폴로지 확장(dilation)을 통해 그림 7(e)와 같은 자막 후보 영역을 얻는다. 그림 7(f)는 자막 후보 영역에 존재하는 에지를 보이고 있는데, 이 에지 중 씨앗점을 포함하는 에지가 자막의 획이 될 가능성이 높다. 그림 7(g)는 자막 후보 영역의 에지 중 씨앗점을 포함하는 에지를 보이고 있다.

자막 위치 찾기를 위한 프로젝션이 효과적으로 수행될 수 있도록 씨앗점을 포함하는 에지의 내부를 채우고 잠음 영역을 제거한 결과가 그림 7(h)와 그림 7(i)이며, 그림 7(j)는 그림 7(i)에 대한 프로젝션을 통해 얻어진 자막 위치의 바운딩박스이다. 그림 7(k)는 바운딩박스 영역에 존재하는 씨앗점의 밝기 평균을 이용하여 이진화를 수행한 결과를 보이고 있는데, “한국”이라는 자막과 “FIFA 랭킹 40위” 라는 자막의 색상이 서로 다른 것을 알 수 있다. 그림 7(l)는 그림 7(k)에 color polarity 기법을 적용하여 “한국”이라는 자막의 색상이 흰색으로 반전된 모습을 보이고 있다. 이때 ‘한’, ‘국’ 주변의 밝은 사각형 테두리는 바운딩박스의 확장된 외곽 영역으로써, 이 영역은 color polarity 기법이 적용되지 않는 부분이다. 이 확장된 영역의 밝은 부분은 dampoint labeling 기법을 통해 제거 된다. 그림 7(m)은 dampoint labeling된 결과를 보이고 있으며, 그림 7(n)은 기하학적 검증을 통해 최종 결과를

도출한 모습을 보이고 있다

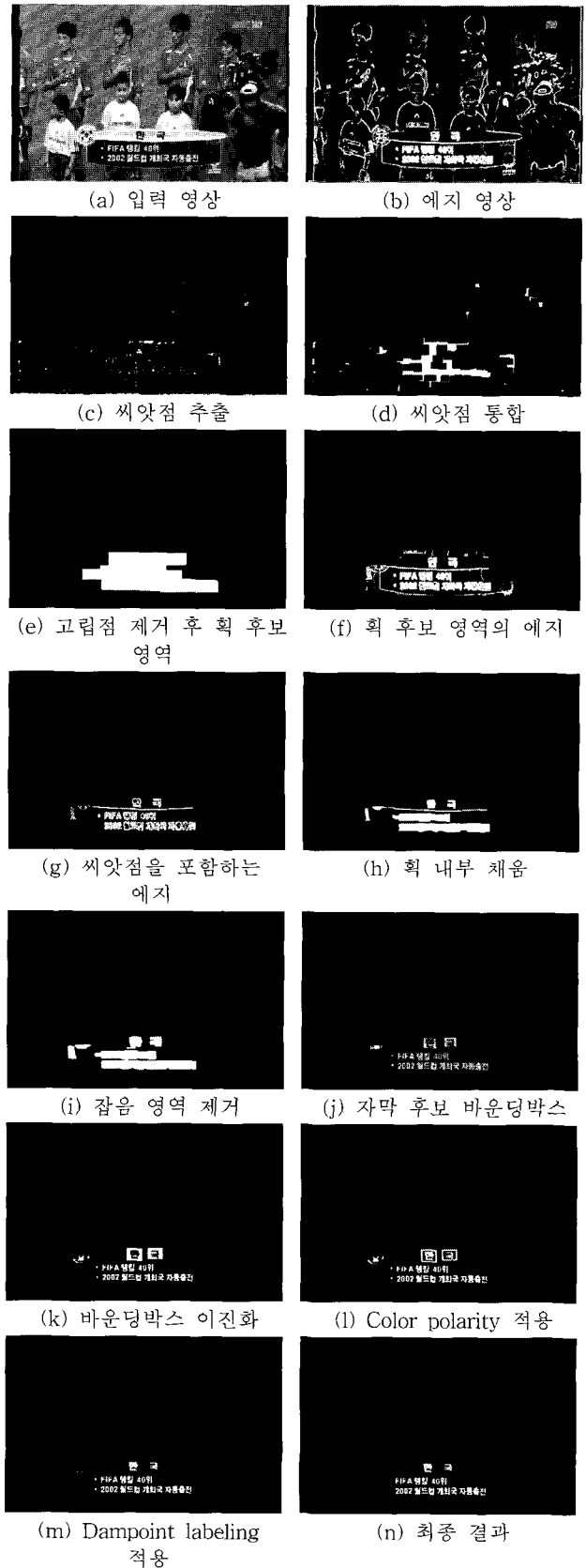


그림 7. 각 단계별 실험 결과.
Fig. 7. Experimental results of each processes.

뉴스, 스포츠, 뮤직 비디오를 등에 존재하는 다양한 크기와 색상을 갖는 4913개의 글자가 존재하는 204개의 프레임에 적용하여 precision과 recall에 의한 성능을 평가한 결과 recall은 94.9%, precision은 95.4% 이었다. 표 1은 실험 결과에 대한 자세한 내용을 보이고 있다.

표 1. 실험 결과 요약.

Table 1. Summarization of experimental results.

장르	뉴스	스포츠	뮤직 비디오	합계
프레임 수	95	67	42	204
자막 수	3131	1411	371	4913
Recall	0.9531	0.9348	0.9677	0.9489
Precision	0.9497	0.9713	0.9229	0.9536

7. 결 론

본 논문에서는 디지털 비디오에서 자막을 추출하기 위한 획기적인 알고리즘을 제안하였다. 제안된 알고리즘은 획의 에지 방향성을 고려하여 씨앗점을 추출하고, 씨앗점들에 대한 모폴로지 연산을 통해 획 후보 영역을 찾은 다음, 획 후보 영역에 존재하는 에지에 대한 프로젝션을 통해 자막 후보 영역의 바운딩박스를 얻는다. 자막 후보 영역 바운딩박스에 존재하는 씨앗점의 밝기 평균을 이용하여 이진화를 수행하고, 이진화된 자막 영역에 대해 color polarity 방법과 dampoint labeling 방법을 이용하여 자막을 분리한 다음, 마지막으로 자막 영역에 대한 기하학적 검증을 통해 최종 결과를 도출하였다. 제안된 알고리즘은 자막의 위치, 색상, 배경, 경계선 등에 관한 별도의 사전 정보 없이 복잡한 배경으로부터 자막을 강건하게 추출할 수 있으며, 이는 실험을 통해 확인되었다.

향후 다중 프레임을 이용한 자막 추출 기법의 개선 및 자막 크기에 무관하게 자막을 추출하기 위한 연구가 필요하다.

참 고 문 헌

[1] K. Jung, K. Kim, A. Jain, "Text Information extraction in images and video: a survey," Pattern Recognition, vol. 37, pp. 977-997, May 2004.

[2] M. R. Lyu, J. Song, and M. Cai, "A Comprehensive Method for Multilingual Video Text Detection, Localization, and Extraction," IEEE Trans. on CSVT., vol. 15n no 2, Feb. 2005.

[3] J. Kim, Y. Moon, "Extraction of Text Regions and Recognition of Characters from Video Inputs," LNCS-Advances in Multimedia Information Processing, vol. 2532, pp. 775-782, Dec. 2002.

[4] R. Lienhart and A. Wernicke, "Localizing and Segmenting Text in Images and Videos," IEEE Trans. on Circuits and Systems for Video technology, vol. 12, no. 4, pp. 256-268, Apr. 2002.

[5] Y. Hasan and L. Karam, "Morphological Text Extraction from Images," IEEE Trans. on Image Processing, vol. 9, no. 11, Nov. 2000.

[6] C. W. Lee, K. Jung, H. J. Kim, "Automatic text

detection and removal in video sequences," Pattern Recognition Letters, vol. 24, pp. 2607-2623, Nov. 2003.

[7] H. Li, D. Doerman, and O. Kia, "Automatic text detection and tracking in digital video," IEEE Trans. on Image Processing, vol. 9, no. 1, pp. 147-156, Jan. 2000.

[8] O. Shiku, Y. Xiao, H. Yan, "Extraction of character patterns in different styles and orientations from natural scene images," Proc. Of 2004 Int. Symp. on Intelligent Multimedia, Video and Speech Processing, pp. 719-722, Oct. 2004.

[9] A. Jian and S. Bhattacharjee, "Text segmentation using gabor filters for automatic document processing," Machine Vis. App., vol. 5, pp. 169-184, 1992.

[10] V. Wu, R. Manmatha, and E. Riseman, "Textfinder: An automatic system to detect and recognize text in images," IEEE Trans. on Pattern Analysis and Machine Intelligent, vol. 21, no. 11, pp. 1224-1229, Nov. 1999.

[11] A. Jain and B. Yu, "Automatic text location in images and video frames," Pattern Recognition, vol. 31, no. 12, pp. 2055-2076, 1998.

[12] M. Cai, J. Song, and M. Lyu, "A new approach for video text detection," Proc. of Int. Conf. on Image Process, pp. 117-120, Sep. 2002.

[13] A. Wernicke and R. Lienhart, "On the segmentation of text in videos," Proc. of IEEE Int. Conf. on Multimedia Expo, vol. 3, pp. 1511-1514, Jul. 2000.

[14] S. Antani, D. Crandall, and R. Kasturi, "Robust extraction of text in video," 15th Int. Conf. on Pattern Recognition, vol. 1, pp. 831-834, Sep. 2001.

[15] R. C. Gonzalez, R. E. Woods, Digital Image Processing, 2nd edition, Prentice Hall, New Jersey, 2001.

[16] J. Song, M. Cai, M. R. Lyu, "A Robust Statistic Method for Classifying Color Polarity of Video Text", Proc. of 2003 International Conference on Multimedia and Expo, vol. 2, pp. II-385-8, Jul. 2003.

저 자 소개



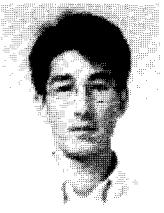
정종면(Jong-Myeon Jeong)

1992년 : 한양대학교 전자계산학과 공학사
1994년 : 한양대학교 대학원 전자계산학과
공학석사
2001년 : 한양대학교 대학원 전자계산학과
공학박사
2001년~2004년 : 한국전자통신연구원 선임
연구원

2004년~현재 : 목포해양대학교 해양전자통신공학부 조교수

관심분야 : 컴퓨터비전, 패턴인식, 멀티미디어 시스템, 대화
형 방송 등

Phone : 061-240-7314
Fax : 061-240-7283
E-mail : jmjeong@mmu.ac.kr



차지훈(Jihun Cha)

1993년 : 명지대학교 전자계산학과 공학사
1996년 : 플로리다공과대학교 전자계산학
과 공학석사
2002년 : 플로리다공과대학교 전자계산학
과 공학박사
2003년~현재 : 한국전자통신연구원 선임
연구원

관심분야 : 대화형 및 이동방송 시스템, 리치미디어, 패턴인식

Phone : 042-860-6305
Fax : 042-860-5479
E-mail : jihun@etri.re.kr



김규현(Kyuheon Kim)

1989년 : 한양대학교 전자공학과 공학사
1996년 : 영국 University of Newcastle
upon Tyne 전기전자공학과 공
학박사
1996년~1997년 : 영국 University of
Sheffield, Reserach Fellow
1997년~2006 한국전자통신연구원 대화형미디어연구팀장
2006년~현재 경희대학교 전자정보대학 부교수

관심분야 : 영상처리, 멀티미디어, 디지털 대화형 방송
Phone : 031-201-3810
Fax : 031-203-0146
E-mail : kyuheonkim@khu.ac.kr