

영화 비디오 자막 추출 및 추출된 자막 이미지 향상 방법

(Methods for Video Caption Extraction and Extracted Caption Image Enhancement)

김 소 명 [†] 곽 상 신 ^{**} 최 영 우 ^{***} 정 규 식 ^{****}
(Somyung Kim) (Sangshin Kwak) (Yeongwoo Choi) (Kyusik Chung)

요 약 디지털 비디오 영상을 효과적으로 색인하고 검색하기 위해서 비디오의 내용을 함축적으로 표현하고 있는 비디오 자막을 추출하여 인식하는 연구가 필요하다. 본 논문에서는 압축되지 않은 비디오 영화 영상에 인위적으로 삽입한 한글 및 영어 자막을 대상으로 자막 영역을 추출하고, 추출된 자막 이미지를 향상시키는 방법을 제안한다. 제안한 방법의 특징은 동일한 내용의 자막을 갖는 프레임들의 위치를 자동으로 찾아서 동일 자막 프레임들을 다중 결합하여 배경에 포함되어 있는 잡영의 일부 또는 전부를 우선 제거한다. 또한, 이 결과 이미지에 해상도 증대, 히스토그램 평활화, 획 기반 이진화, 스무딩의 이미지 향상 방법을 단계적으로 적용하여 인식 가능한 수준의 이미지로 향상시킨다. 제안한 방법을 비디오 영상에 적용하여 동일한 내용의 자막 그룹 단위로 자막 이미지를 추출하는 것이 가능해졌으며, 잡영이 제거되고 복잡한 자소의 획이 보존된 자막 이미지를 추출할 수 있었다. 동일한 내용의 자막 프레임의 시작 및 끝 위치를 파악하는 것은 비디오 영상의 색인과 검색에 유용하게 활용될 수 있다. 한글 및 영어 비디오 영화 자막에 제안한 방법을 적용하여 향상된 문자 인식 결과를 얻었다.

키워드 : 비디오 색인, 자막 추출, 이미지 향상, 다중 결합, 문자 인식

Abstract For an efficient indexing and retrieval of digital video data, research on video caption extraction and recognition is required. This paper proposes methods for extracting artificial captions from video data and enhancing their image quality for an accurate Hangul and English character recognition. In the proposed methods, we first find locations of beginning and ending frames of the same caption contents and combine those multiple frames in each group by logical operation to remove background noises. During this process an evaluation is performed for detecting the integrated results with different caption images. After the multiple video frames are integrated, four different image enhancement techniques are applied to the image: resolution enhancement, contrast enhancement, stroke-based binarization, and morphological smoothing operations. By applying these operations to the video frames we can even improve the image quality of phonemes with complex strokes. Finding the beginning and ending locations of the frames with the same caption contents can be effectively used for the digital video indexing and browsing. We have tested the proposed methods with the video caption images containing both Hangul and English characters from cinema, and obtained the improved results of the character recognition.

Key words : Video indexing, Caption extraction, Image enhancement, Multi-frame integration, Character recognition

본 연구는 과학기술부 뇌신경정보화사업의 지원을 받아 수행되었음 (M1-0107-00-0009).

[†] 비 회 원 : LG전자 디지털 TV 연구소 연구원
somyung@lge.com

^{**} 비 회 원 : OrCom 주임 연구원
gguggu@orcom.co.kr

^{***} 정 회 원 : 숙명여자대학교 정보과학부 교수
ywchoi@sookmyung.ac.kr

^{****} 종신회원 : 숭실대학교 정보통신전자공학부 교수
kchung@q.soongsil.ac.kr

논문접수 : 2001년 1월 27일
심사완료 : 2002년 1월 21일

1. 서론

컴퓨터, 압축 기술, 저장 매체, 고속 통신기술의 발달로 인한 멀티미디어 정보화 시대에서 대표적인 멀티미디어 중의 하나인 디지털 비디오는 교육, 오락 등을 비롯한 각종 응용 분야에서 중요한 부분을 차지하고 있다. 이에 따라 방대한 양의 디지털 비디오 영상을 빠르고 효과적으로 검색하기 위해 비디오 자막의 추출, 인식 및 색인에 관한 연구가 계속 진행되고 있다.

비디오 영상에 포함되어 있는 자막은 비디오의 내용을 함축적으로 표현하고 있기 때문에 이 자막을 정확하게 인식할 수 있다면 비디오의 색인 및 검색에 중요하게 사용될 수 있다. 또한 비디오 자막은 동/정지 영상과 음성, 음향 정보에서 표현하고 있지 않는 내용도 포함하는 경우가 있어서 그 효용성이 크다. 그러나 그림 1과 같이 비디오 자막 이미지의 해상도가 낮고 복잡한 배경이 포함될 수 있기 때문에 비디오 자막을 인식하는 것은 아주 어려운 일이다.



그림 1 낮은 해상도와 복잡한 배경을 포함한 자막 이미지의 예

비디오 자막을 정확하게 인식하기 위해서는 자막 영역을 추출하는 과정과 자막 이미지의 해상도를 증대시키고, 복잡한 배경을 제거하는 이미지 향상 과정이 필수적이다. 자막 추출에 관련된 연구를 살펴보면 압축 영상에서 압축을 풀지 않고 처리를 하는 연구와[1-3] 비압축 영상 또는 압축 영상에서 압축을 풀어서 처리를 하는 연구로[4-8] 나누어 생각할 수 있다. 압축 영상에서 압축을 풀지 않은 상태에서 자막 영역을 추출하는 경우에는 주로 I(Intracoded) 프레임의 AC 계수를 이용했으며, 비압축 영상 및 압축을 푼 영상에 대해서는 연결 요소나 명암도의 차이, 그리고 각 실험 영상의 사전 지식 정보 등을 이용하였다.

이미지 향상에 관련된 연구를 살펴보면 선형 보간법이나 Shannon 보간법을 이용하여 영상의 해상도를 증대시킨 후 적응적 이진화를 적용하거나 다중 결합(Multiple-frame Integration)을 통해 복잡한 배경을 제거하는 방법 등이 제안되었다[8,9]. 그러나 Shannon 보간법은 시간이 많이 걸리는 단점이 있으며, 적응적 이진화 방법을 사용하면 특히 한글 이미지에서 복잡한 획이 쉽게 뭉쳐지는 결과가 발생하는 단점이 있다. 이러한 연구들에서 부족한 점은 다중 결합에 필요한 동일한 내용의 자막을 갖는 프레임의 위치를 자동으로 판별하는 과

정이 생략된 것과 추출된 이미지를 개선하는 과정에서 한글과 같이 복잡한 자소로 구성된 문자들의 획이 쉽게 뭉쳐지는 것을 방지하는 방법에 대한 언급이 없는 것이다. 다중 결합은 자막 주변의 배경 잡영을 제거할 수 있는 효율적인 방법이지만, 이것을 수행하기 위해서는 동일한 내용의 자막 프레임들을 자동으로 정확히 판별하는 과정이 필수적이다.

본 논문에서는 위에서 지적한 문제점들을 개선하기 위해서 우선 동일한 내용의 자막 프레임들의 위치를 자동으로 판단하는 방법을 제안하고, 이 방법에 따라 프레임들을 그룹화 시킨 후 같은 그룹 내의 프레임들을 다중 결합하여 그룹 단위로 배경 잡영이 제거된 자막 이미지를 만든다. 또한 한글의 복잡한 자소에서 쉽게 발생하는 획의 뭉침을 개선하기 위한 이미지 향상 방법들도 제안한다. 동일한 내용의 자막 프레임의 시작 및 끝 위치를 정확하게 파악하는 것은 내용 기반 색인에 도움을 줄 수 있다. 제안한 방법들은 실험 영상의 종류 및 특성에 따라 선택적으로 적용하여 처리 시간을 단축시킬 수 있다. 제안하는 방법은 그림 2와 같이 자막 추출, 동일 자막 프레임 판별 및 자막 이미지 향상의 세 단계로 진행된다.

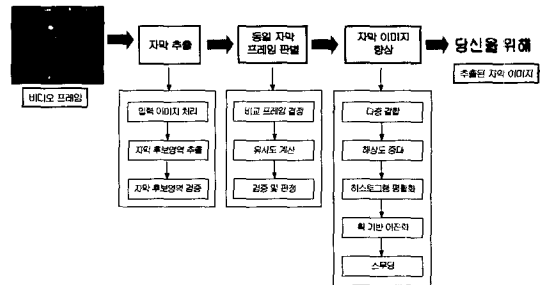


그림 2 제안한 방법의 수행 단계

본 논문의 구성으로서 2절에서는 관련된 기존 연구를 소개하며, 3절과 4절에서는 제안하는 방법을 구체적으로 기술한다. 5절에는 실험 환경 및 결과를 실었으며, 6절에서 결론을 맺는다.

2. 관련 연구

관련 연구를 살펴보면 크게 압축 영상에서 압축을 풀지 않은 상태에서 처리하는 연구와 비압축 영상 또는 압축 영상에서 압축을 풀어서 처리를 하는 연구로 나누어 생각할 수 있다. 먼저 압축 영상에서 압축을 풀지 않고 자막 영역을 추출하는 방법을 살펴보면, I 프레임에서 수직 예지와 관련된 AC 계수들의 합이 일정한 값

이상인 블록들을 자막 후보 블록으로 결정하고, 수평 에지와 관련된 AC 계수들의 합과 모폴로지 등을 이용하여 최종적으로 자막 영역을 결정하는 방법이 제안되었다[1]. 이 방법은 이미지의 해상도에는 크게 영향을 받지 않는 장점이 있으나, 문자의 크기나 문자 사이의 간격이 너무 큰 경우에는 자막 영역으로 찾아내지 못하는 단점이 있다. 이성환 등의[2] 연구에서는 I 프레임의 지그재그 형태의 스캔 배열에서 DC 계수를 제외한 처음 다섯 개의 AC 계수의 합이 일정한 값 이상인 블록들을 후보 블록으로 결정하고, DC 계수와 P(Predictive) 프레임에서의 매크로 블록 타입을 이용하여 검증하였다. 이 방법은 P 프레임을 이용함으로써 배경이 잘못 추출된 경우를 줄일 수 있었지만, 배경이 움직이는 자막 영역도 제거되는 단점이 있다.

비압축 영상 또는 압축을 풀어서 처리하는 연구를 살펴보면, Bolle 등의[4] 연구에서는 명도 값의 차이가 임계값 이하인 화소들에 대해 외곽 추적(Contour traversing)을 통해 라벨링을 수행하고, 사전 정보를 이용하여 각 컴포넌트들을 묶음(Clustering)으로써 자막 영역을 결정하였다. 그리고 개별적인 프레임에서 찾지 못한 자막 영역이나 문자 부분을 찾아내기 위해서 인접한 다섯 개의 프레임들과 추출 영역을 비교한다. 위의 방법은 인접한 프레임들의 추출 영역까지 고려함으로써 자막 영역을 보다 정확하게 찾지만, 인접한 프레임들이 서로 다른 자막을 포함하고 있는 경우에는 정확하게 추출하는 것이 어려운 단점도 있다. 자막을 하나의 특정한 텍스처로 설정하고 가우시안 필터를 적용하는 방법이 제안되었다[10

]. 필터를 적용한 후에 원래 이미지의 각 화소들을 특징 벡터로 표현하고, 적절한 클러스터링 방법을 이용하여 자막에 해당하는 영역을 추출하였다. 이와 같이 필터를 적용하는 방법은 다양한 폰트 종류와 크기에 대해서도 적용할 수 있지만 시간이 오래 걸리는 단점이 있다.

전병태 등은[11] 명암도 차, 획의 폭과 문자간의 거리에 대한 특징점을 추출하고, 점으로부터 자막 영역의 조건을 만족하는 점들을 연결하여 선을 만들고, 선으로부터 면으로 확장하여 영역을 추출하였다. 이 방법은 문자에 대한 사전 지식 없이도 문자 영역을 추출할 수 있지만, 점, 선, 면의 각 레벨에서 추출된 영역들이 일치하지 않는 경우에 어떠한 영역을 선택할 것인지 결정하는 방법에 있어서 보완이 필요하다. 전병태 등의[12] 연구에서는 자막의 명암 파형의 특성을 분석하여 고주파 부분만을 후보 영역으로 추출하였으며, 이성환 등의[7] 연구와 Smith 등의[8] 연구에서는 자막과 배경은 높은 대조를 이룬다는 특징을 이용하여 자막에 대한 에지를 찾아내고 에지 성분이 조밀하게 분포되어 있는 영역을 자막 영역으로 간주하였다.

표 1은 위에 언급한 관련 연구들과 제안한 방법을 자막 영역의 추출 방법과 동일 자막 판단 여부에 대하여 비교한 것이다. 표 1에서 볼 수 있듯이, 대부분의 논문에서 자막 후보 영역의 추출과 검증에 대한 방법들은 제안하였으나, 동일 자막 프레임의 판단에 대한 방법에 대해서는 언급하고 있지 않다. 동일 자막 프레임들을 판단하는 것은 비디오의 자막을 추출하는 많은 연구들이 효과적인 정보 검색을 수행하기 위해서 반드시 필요한

표 1 관련 연구에 대한 비교

구분	참고문헌	영역 추출 방법	검증 방법	동일 자막 프레임 판단 과정
압축 영상	[1]	I 프레임의 수직 에지에 대한 AC 계수의 합	수평 에지의 AC 계수의 합, 후보 영역 크기	없음
	[2]	I 프레임의 처음 5개의 AC 계수의 합	DC 계수, P 프레임의 매크로 블록의 타입	없음
비압축 영상	[4]	명도 값 이용, 외곽 추적	문자 영역의 특성, 인접 프레임의 추출영역과 비교	없음
	[7]	자막 영역과 배경의 명도차	YIQ 칼라 마스크, RLS 적용	없음
	[8]		후보 영역의 수직, 수평 비율, 후보 영역의 크기 등	없음
	[10]	가우시안(Gaussian) 필터	후보 영역의 수직·수평 비율, 후보 영역의 크기 등	없음
	[11]	특징점, 점·선·면 확장법	후보 영역의 수직·수평 비율, 후보 영역의 크기 등	없음
	[12]	명암 파형 특성	후보 영역의 크기 및 개수	없음
제안방법	명도 값 이용(에지 추출), RLS 적용	에지 밀도 값, 대략적인 사전 위치 정보	있음	

과정이다. 따라서, 본 연구에서는 동일 자막 프레임들을 판단하는 방법을 제안하는 것에 중점을 두었다.

이미지 향상 관점에서의 연구를 살펴보면 선형 보간법을 이용하여 영상의 해상도를 증대시키고 다중 결합을 수행하여 복잡한 배경을 제거하는 방법이 Smith 등에[8] 의하여 제안되었다. 이 연구에서 사용한 선형 보간법은 단순하고 빠른 장점이 있지만, 획의 경계가 모호해지거나 획의 일부만이 강조되는 단점이 있다. Doermann 등은[9] Shannon 보간법과 적응적 이진화를 적용하여 이미지를 향상시켰다. 그러나 Shannon 보간법은 이미지의 크기가 커질수록 처리 시간이 급격히 증가하는 단점이 있으며, 간단한 적응적 이진화 방법을 한글 이미지에 적용하면 복잡한 획이 쉽게 뭉쳐져서 글자를 식별하기 어렵게 만드는 단점이 있다.

본 연구에서는 위에서 살펴본 것과 같이 기존의 연구들에서 언급하지 않은 동일한 자막 프레임의 위치를 자동으로 판단하는 방법을 제안하여 내용 기반 비디오 색인 및 촬영 제거를 위한 다중 결합에 도움을 주고자 한다. 또한, 비디오 영상에 포함되어 있는 한글 이미지와 같이 획이 쉽게 뭉쳐지는 것을 방지할 수 있는 이미지 개선 방법들도 제안하고자 한다.

본 연구는 비디오 영상에 인위적으로(Artificial) 삽입한 자막 텍스트의 추출 및 인식에 있지만, 다른 한편으로는 이미지 또는 동영상에 자연적으로(Natural) 포함된 장면(Scene) 텍스트 추출에 관한 연구들도 있다[13-16]. 장면 텍스트들은 아무런 제약 없이 이미지 상에 위치하며 다양한 종류의 폰트, 크기, 간격, 배열, 방향, 색, 무늬, 원근감 등을 갖고 나타나며, 이미지가 취득되는 순간의 조명과 기상 여건에 따라 다양한 질(質)의 이미지가 획득되어 추출을 더욱 어렵게 만든다. 장면 텍스트 추출의 연구 결과는 로봇 자동 주행시스템, 시각장애인 보행 안내시스템, 공장 자동화, 이미지/비디오 색인 및 검색 등에 활용될 수 있다. 추후 연구에서 고려할 예정이다.

3. 자막 추출 및 동일 자막 프레임 판단

본 절에서는 자막 영역을 추출하는 방법과 동일한 자막 프레임들을 판단하는 방법을 기술한다.

3.1 자막 추출

자막 영역을 추출하기 위해서 칼라 이미지를 먼저 명도 이미지로 변환한 후 에지 이미지를 구한다. 에지 이미지의 자막 영역을 강조하기 위해서 수평 및 수직 방향으로 RLS(Run-length Smearing)를 적용한다. RLS가 적용된 에지 이미지로부터 수직 및 수평 방향의 히스토그램을 분석하여, 에지 분포가 조밀한 영역을 자막이 존재할 수 있는 후보 영역으로 설정한다. 비디오 영상의 자막에 관련된 사전 정보인 자막의 위치와 자막의 에지 밀도를 이용하여 자막 후보 영역을 검증한다. 추출된 자막 영역의 위치, 크기, 에지 밀도, 프레임 내의 자막 개수에 관한 정보를 저장한다.

3.1.1 입력 이미지 처리

비디오 영상을 처리하기 위한 첫 번째 단계로서 칼라 이미지를 명도 이미지로 변환시킨다. 명도 이미지는 비디오 영상 처리에 흔히 사용되고 있으며, 비교적 빠르고 쉽게 처리되는 장점이 있다. 일반적으로 비디오 자막은 밝은 명도 값과 배경과의 높은 대비를 가지며 배경과 뚜렷이 구분되는 외곽선을 갖기도 한다. 본 논문에서도 이러한 일반적인 특징들을 활용하여 자막이 포함되어 있는 프레임을 결정하고 자막 후보 영역을 찾는다. Sobel 연산자[17]를 사용하여 에지 이미지를 생성하며, 수평 및 수직 방향의 RLS를 적용하여 자막 영역을 강조시킨다. 에지를 찾는 과정에서의 임계값(Threshold)은 실험을 통해서 가장 적절한 값으로 설정하였다. 그림 3은 위의 과정이 수행되면서 생성된 이미지들로서 (b)는 원영상 (a)로부터 얻은 에지 이미지이며, (c)는 (b)에 RLS를 적용하여 생성된 이미지로서 자막 영역이 강조된 것을 확인할 수 있다.

3.1.2 자막 후보 영역 추출

RLS가 적용된 에지 이미지에서 자막 영역은 주변 영역에 비해 수평 방향으로 높은 에지 분포를 갖기 때문

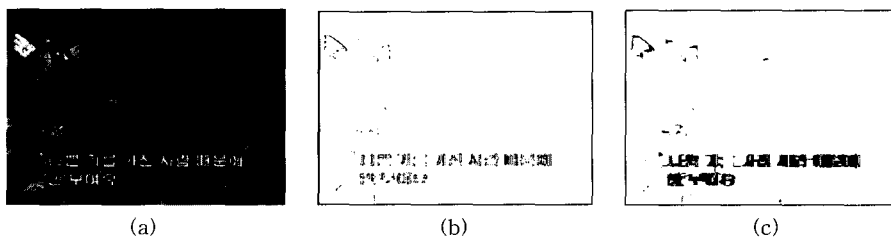


그림 3 입력 이미지 처리 : (a) 원영상, (b) 에지 추출, (c) RLS가 적용된 결과

에 RLS 결과의 수직 및 수평 에지 분포를 분석하여 자막 영역을 결정한다. 먼저 자막의 수직 영역을 결정하기 위해서 에지 이미지의 수평 에지의 분포가 조밀한 영역을 찾는다. 그림 4(a)는 수평 에지 분포의 프로파일을 보여주며 결정된 자막의 상하 후보 영역을 CR1과 CR2의 사각형으로 표시한다. 또한, 자막의 좌우 영역을 결정하기 위해서 그림 4(b)와 같이 비디오 자막의 한 문자 크기 정도의 윈도우와 윈도우 내에서의 수직 에지 분포 프로파일을 분석한다. 윈도우를 설정하여 사용하는 이유는 단순히 수직 에지 값이 존재하는 영역을 자막의 좌우 영역으로 결정하면, 자막 주변에 배경으로 인해서 발생한 에지가 존재하는 경우에 실제 자막보다 더 넓은 영역이 자막 후보 영역으로 설정되기 때문이다. 자막의 수직 후보 영역인 CR에서 윈도우를 왼쪽에서 오른쪽으로 옮겨가며 윈도우내에 분포된 수직 에지의 개수의 합이 일정한 값 이상인 영역만을 자막 후보 영역의 좌우 경계로 설정한다.

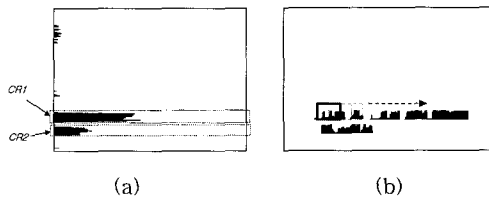


그림 4 (a) 에지의 수평 분포 분석에 의한 자막의 상하 영역 결정, (b) 에지의 수직 분포 분석에 자막의 좌우 영역 결정

3.1.3 자막 후보 영역 검증

수직 및 수평 에지 프로파일만을 분석하여 자막 후보 영역을 결정하면 자막에 배경이 포함되어 추출되거나 배경만 추출되는 경우도 발생한다. 따라서 추출된 자막 후보 영역에 대한 검증이 필요하다. 본 논문에서 사용한 검증 특징은 비디오 영상에 존재하는 자막의 사전 위치 정보와 자막 영역의 에지 밀도 값이다. 본 연구의 처리 대상이 영화 비디오이기 때문에 전체 프레임 높이의 1/3 이하인 위치에서 추출된 자막 후보 영역만을 고려하였으며, 후보 영역의 에지 밀도 값이 일정한 값 이상인 경우만을 자막 후보 영역으로 판정한다. 자막 영역이 추출되는 과정에서 하나의 자막이 두 영역으로 분할되어 추출되는 경우도 발생하기 때문에 같은 높이에 위치하며 좌우로 인접한 후보 영역들은 결합시킨다. 최종 결정된 자막 후보 영역의 위치, 크기, 에지 밀도, 자막 개수에 관한 정보를 저장한다.

3.2 동일 자막 프레임 판별

비디오 영상에 포함된 동일한 내용의 자막은 선형적으로 움직이는 자막이 아니라면 연속적인 프레임들의 동일한 위치에 나타나며, 자막의 크기와 폰트는 일정하며 명도 값은 거의 비슷한 특징들이 있다. 본 절에서는 이러한 정보들을 이용하여 동일한 자막 프레임들을 자동으로 결정하는 과정을 설명한다.

3.2.1 비교 프레임 결정

비교 프레임의 결정은 현재 프레임이 이전 프레임과 동일한 자막을 가질 가능성이 있는지를 판단하는 과정이다. 즉 동일한 내용의 자막을 가질 가능성이 없을 정도로 두 프레임 사이의 시간 간격이 크다면 미리 비교 대상에서 제외시킴으로써 인접한 프레임과의 비교에 소요되는 처리 시간을 줄인다. 자막이 추출된 두 프레임 사이에 자막이 추출되지 않은 프레임 개수가 정해진 값보다 작다면, 동일한 자막이 연속적으로 나타나지만 중간에 일시적으로 자막 추출이 실패한 경우로 생각하여 비교 대상 프레임으로 결정한다.

두 프레임이 비교 대상 프레임으로 결정되면 정의된 유사도(Similarity) 계산 방법에 의하여 동일 자막 프레임 여부를 판단한다. 반면 자막이 추출된 두 프레임 사이에 자막이 추출되지 않은 프레임의 개수가 정해진 값보다 크면 각 프레임은 서로 다른 자막을 가진 프레임으로 판단하여 비교 대상 프레임에서 제외시킨다. 임계값은 실험 대상 영상으로부터 최적인 값으로 설정하였다.

3.2.2 유사도 계산

현재 자막이 추출된 프레임 F_i 가 바로 이전에 자막이 추출된 프레임 F_{i-1} 과 동일한 자막을 포함하고 있는지를 판별하기 위해서 두 프레임간의 유사도를 계산한다. 유사도 S 는 현재 프레임에서 추출된 자막 영역과 비교되는 이전 프레임의 추출된 자막 영역들의 영역 중첩 비율로서 식 (1)과 같이 정의한다.

$$S(F_i, F_{i-1}) = \frac{1}{N} \sum_{i=1}^N \left\{ \left(\frac{OS_i}{PS_i + CS_i} \right) \times 2 \right\} \quad (1)$$

식 (1)에서 PS 는 비교되는 이전 프레임의 자막 영역의 크기이며, CS 는 현재 프레임의 자막 영역의 크기이다. OS 는 두 자막 영역 PS 와 CS 가 중첩되는 영역의 크기이며, N 은 자막의 개수로서 현재 프레임과 이전 프레임을 비교하여 큰 자막의 개수로 정하며, i 는 한 프레임 내에서 부여된 자막의 일련 번호이다. 식 (1)의 유사도 S 는 프레임에서 추출된 자막 영역의 개수와 자막 영역의 중첩 정도에 따라 결정된다. 유사도가 정해진 값보다 크면 현재 프레임과 이전 프레임에 동일한 자막이 있다고 판단하여 동일한 라벨(Label)을 부여한다.

이 과정에서 바로 이전 프레임과의 비교만으로 동일 자막 여부를 판단하면 자막 주변 영역에 포함되는 잡영에 의하여 일시적으로 자막 영역이 다르게 추출될 수 있기 때문에, 동일한 내용의 자막 프레임들에게도 서로 다른 라벨이 부여되는 문제점이 발생한다. 따라서 그림 5와 같이 자막이 추출된 프레임 F_{t-2} 와의 유사도도 검증한다. 우선 F_t 와 F_{t-1} 프레임을 비교한 후 유사도 값이 낮아서 동일하지 않다고 판단된 경우에 F_t 와 F_{t-2} 프레임을 비교한다. F_t 와 F_{t-2} 의 유사도 값이 높으면 F_t , F_{t-1} , F_{t-2} 프레임들에 대하여 모두 동일한 라벨을 부여한다. 여기서도 3.2.1절에서 설명한 것과 같이 자막이 추출된 프레임들 사이에 자막이 추출되지 않은 프레임의 개수 정보인 count1과 count2에 대한 조건이 만족되어야만 F_t 와 F_{t-1} 또는 F_t 와 F_{t-2} 에 대한 유사도가 각각 측정된다. 그림 6은 위의 방법을 순서도로 정리한 것이다.

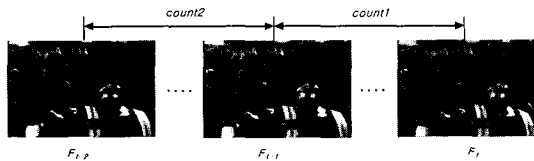


그림 5 동일 자막 판별 및 라벨링

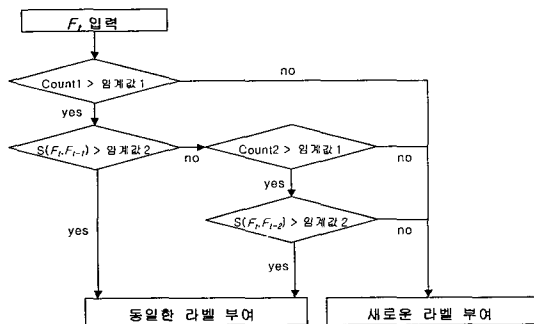


그림 6 유사도를 이용한 동일 자막 프레임 판별

3.2.3 검증 및 판정

추출된 자막 영역은 자막 주위의 배경으로 인해서 자막의 실제 크기보다 작거나 크게 추출되는 경우가 빈번하게 발생한다. 따라서 자막 영역들이 중첩되는 정도만을 고려하여 동일 자막 여부를 판단하면 서로 다른 내용의 자막 프레임들이 동일한 자막 프레임으로 판단되거나 동일한 자막 프레임들이 서로 다른 자막 프레임으로 판단되기도 한다. 이와 같은 오류를 줄이기 위한 검

증 과정이 필요하며 검증에 사용된 특징은 자막 영역의 에지 개수이다. 즉 비교되는 두 프레임의 중첩된 영역에서의 에지 개수가 비슷해야만 동일한 자막으로 최종 판정한다.

검증을 마치고 최종 판정된 자막 그룹들에는 적은 수의 프레임만으로 구성된 그룹들도 있다. 이러한 그룹내의 프레임들은 대부분 자막은 없지만 에지가 조밀하게 분포된 배경 영역을 일시적으로 자막 영역으로 잘못 판단하여 추출한 경우이다. 따라서 이러한 그룹은 자막이 없는 프레임에서 잘못 추출한 것으로 판단하여 무시한다. 동일 자막 프레임의 판정 결과는 표 2와 같이 다섯 종류로 분류한다. 유형 (I), (II), (III)은 동일한 자막 프레임들로 구성되기 때문에 다중 결합을 수행하여 잡영의 일부 또는 전부를 제거할 수 있지만, 유형 (IV), (V)는 서로 다른 자막 프레임들에 동일한 라벨이 부여된 경우로서 다중 결합을 수행하면 자막 이미지가 훼손되어 인식이 더욱 어렵게 된다. 따라서 유형 (IV), (V)를 판단하는 재평가 과정과 재평가 결과에 따른 재분류 과정이 필요하다.

표 2 동일 자막 프레임의 판정 결과

유형 분류	처 리
(I) 동일한 자막 프레임에 하나의 동일한 라벨이 부여된 경우	재분류 생략
(II) 동일한 자막 프레임들의 일부만 동일한 라벨이 부여된 경우	
(III) 동일한 자막 프레임들에 두 개 이상의 서로 다른 라벨이 부여된 경우	
(IV) 서로 다른 두 개의 자막 그룹에 동일한 라벨이 부여된 경우	재분류 시도
(V) 동일한 라벨이 서로 다른 두 개의 자막 그룹의 일부 프레임들에 부여된 경우	

4. 자막 이미지 향상

추출된 자막을 정확하게 인식하기 위해서는 자막 이미지의 해상도를 증가시키고 복잡한 배경을 제거하는 이미지 향상 과정이 필수적이다. 특히 복잡한 자소의 조합으로 구성된 한글 자막을 정확하게 인식하기 위해서는 더욱 정교한 이미지 향상 방법들이 요구된다. 본 연구에서는 우선 앞 절에서 판정한 동일 자막 프레임들간의 다중 결합을 수행하여 배경에 포함된 잡영의 일부 또는 전부를 제거한 후, 해상도 증대, 히스토그램 평활화, 획 기반 이진화, 스무딩 과정을 순차적으로 적용하여 이미지의 질을 더욱 개선한다. 본 연구에서는 이미지

질의 향상에 초점을 맞추었으며, 이 과정에서 소요되는 처리 시간은 현재 고려하지 않았다. 비디오 영상의 종류 및 특성에 따라 이미지 향상 방법들을 선택적으로 적용함으로써 처리 시간 문제를 추후에 해결하고자 한다.

4.1 다중 결합

비디오 영상의 동일한 자막들은 대부분 연속적인 프레임들의 동일한 위치에 나타나지만 자막이 아닌 배경들은 변하거나 움직이는 경우가 대부분이다. 즉 자막에 해당되는 화소 값은 여러 프레임의 동일한 위치에서 밝은 명도 값을 유지하지만, 배경에 해당되는 화소 값은 같은 위치에서 명도 값이 변하는 경우가 많다. 이러한 특징을 이용하여 동일한 자막 프레임들을 식 (2)와 같이 다중 결합하여 자막 주변에 포함된 배경의 일부 또는 전부를 제거할 수 있다. 다중 결합은 동일 자막 프레임들의 동일한 위치의 화소 값들 중에서 최소 값, 즉 가장 어두운 값을 최종 결과 값으로 선택함으로써 이루어진다. 즉, 한 프레임만을 고려하여 자막 영역을 추출하게 되면 배경에 해당되는 화소 값이 밝은 명도 값인 경우에는 잠영으로 남지만, 동일한 자막 프레임들을 비교하여 화소 값을 결정하면 배경이 움직이는 경우가 많기 때문에 결국 화소 값은 배경 가운데 가장 어두운 값에 해당하는 화소 값으로 설정되어 밝은 배경이 제거된다.

$$MI_i(x, y) = \min(F_m(x, y), F_{m+1}(x, y), \dots, F_n(x, y)) \quad (2)$$

식 (2)에서 $MI_i(x, y)$ 는 i 번째 라벨을 갖는 프레임들의 동일한 위치 x, y 에 대하여 다중 결합된 화소의 결과 값이며, $F_j(x, y)$ 는 j 번째 프레임의 x, y 좌표에서의 화소 값이다. m 과 n 은 각각 i 번째 라벨을 갖는 그룹의 시작과 끝 프레임 번호이다.

동일한 자막 프레임들에 하나의 라벨이 부여된 경우는 그림 7(a) 처럼 다중 결합을 수행하면 배경 잡영이 제거된 결과를 얻을 수 있다. 그러나 서로 다른 자막 프레임에 동일한 라벨이 부여된 경우는 그림 7(b)와 같이 서로 다른 자막들이 겹쳐져서 문자의 획이 훼손되어 인식이 더욱 어려워진다. 이러한 문제점을 해결하기 위해서 본 논문에서는 서로 다른 내용의 자막이 결합되면 밝은 화소의 개수가 급격히 감소한다는 특징을 이용하여 다음과 같은 과정을 수행한다. 다중 결합된 자막 이미지에서 밝은 화소의 개수를 검사하여 그 값이 급격하게 감소되었다고 판단되면, 결합된 프레임들을 다시 한 프레임씩 결합하면서 설정된 자막 영역의 밝은 화소의 개수를 계속 검사한다. 결합 도중에 밝은 화소의 개수가 급격히 변하게 되는 프레임을 새로운 자막의 시작 프레임으로 다시 분류한다. 설정된 시작 프레임을 기준으로 두 그룹으로 나누어 각각 다중 결합을 수행하면 잡영이

제거된 결과 이미지를 얻을 수 있다. 실험 결과는 5절에 있다.



그림 7 다중 결합 결과 : (a) 동일한 자막들이 결합된 경우, (b) 서로 다른 자막들이 결합된 경우

4.2 해상도 증대

표준 MPEG인 경우 프레임 크기는 352×240 정도의 해상도를 가지며 이러한 프레임에 삽입되는 한글 자막은 일반적으로 16×16 픽셀 이하의 문자 크기를 가진다. 또한 CNN 뉴스나 영화 비디오에 삽입되는 영어자막은 8×8 픽셀 보다 작은 문자들도 있다. 이와 같이 저 해상도의 자막을 인식하기 위해서는 이미지의 해상도를 증대하는 과정이 필요하다. 현재 많이 사용되고 있는 선형 보간법을 이용한 해상도 증대 방법은 처리 과정이 간단하며 계산 속도가 빠른 장점이 있지만[9], 문자 획의 경계가 모호해지거나 획의 일부분만이 강조되는 단점도 있다. 선형 보간법보다 성능이 우수한 sinc-함수 보간법 방법은[8] 확대된 이미지에 저대역 필터를 통과시켜서 이미지의 해상도를 증대한다. 이 방법을 사용하면 문자 획의 경계가 보다 선명하게 나타나지만, 이미지의 크기에 따라 처리 시간이 급격히 증가하는 단점이 있다. 본 논문에서는 비교적 처리 과정이 간단하고 처리 시간이 빠른 선형 보간법을 이용하여 해상도를 증대시키며, 이 방법의 단점은 뒤에 적용되는 이미지 향상 과정들을 사용하여 보완하고자 한다.

4.3 히스토그램 평활화

다중 결합을 수행하여 배경의 일부 또는 전부가 제거되기도 하지만 배경이 움직이지 않는 경우에는 다중 결합을 사용하여 배경을 제거할 수 없다. 따라서 히스토그램 평활화 과정을 적용하여 배경을 추가로 제거시키고자 한다. 우선 자막 이미지를 일정한 간격의 작은 영역으로 나누고 각 영역의 명도 값의 분포를 파악하여 상, 하 임계값을 각각 결정한다. 다음으로 각 영역에서 두 임계값 사이에 있는 화소의 명도 값들은 명도 값 0과 255 사이에서 선형적으로 다시 분포시키며, 임계값보다 작거나 큰 값은 명도 값 0과 255로 설정한다.

4.4 획 기반 이진화

본 논문에서는 Zhao 등에[18] 의해 제안된 이진화 방법을 히스토그램 평활화가 적용된 결과 이미지에 적

용한다. 이 방법은 저 해상도의 이미지에서도 “를”, “을”, “똥” 등과 같이 복잡한 유형의 중성을 포함한 한글 자막 이미지에서도 좋은 결과를 보여 준다. 이 방법은 다음과 같이 요약되며 자세한 사항은 참고문헌을 참조하기 바란다.

획 기반 이진화 방법을 간단히 정리하면 우선 추출된 자막 이미지에서 문자 획의 두께 W 를 자동으로 파악한다. 다음으로 처리중인 화소의 명도 값과 일정한 거리만큼 떨어져 있는 4개의 지역 영역을 결정하고 각 영역에서 얻어진 평균 명도 값과 동시에 비교한다. 4개 영역의 선정은 현재 화소로부터 $2W+1$ 만큼의 거리에 위치한 8방향의 지역 영역에서 선택한다. 각 지역 영역의 크기는 실험에 의하여 결정하며, 선택된 4개의 영역은 하나의 방향성 그룹을 표현한다. 각각의 방향성 그룹은 (북서, 북, 남동, 남), (북, 북동, 남, 남서), (북동, 동, 남서, 서)와 (동, 남동, 서, 북서)이다. 각 그룹의 평균 명도 값과 현재 화소의 명도 값과의 차이를 순차적으로 비교하여 이 중 어느 하나의 값 차이가 임계값 이상이면 그 그룹이 표현하는 방향성과 반대 방향에 획이 존재한다고 판단하여 현재의 화소 값을 1로 정한다.

4.5 스무딩(Smoothing)

추출된 자막은 획 사이의 간섭 또는 선형 보간법을 이용한 해상도 증대로 인해서 문자 획에 작은 잡영들이 포함되어 문자 인식의 정확성을 떨어뜨린다. 또한, 문자와 잡영이 접촉되면 자막을 최종 인식하는데 필요한 과정인 문자 분리가 어려워진다. 이러한 문제점들을 해결하기 위하여 모폴로지를 사용하여 이미지를 스무딩시킨다. 모폴로지에 사용된 Structure element는 그림 8과 같으며 Dilation과 Erosion을 각각 한번씩 실시한 Opening을 적용한다.

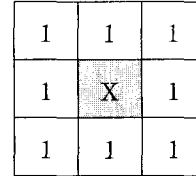


그림 8 모폴로지에 사용된 Structure Element

5. 실험 및 결과

Pentium III 600MHz PC에서 Visual C++ 6.0을 사용하여 제안한 방법들을 구현하였으며, 한글 비디오 영상 두 개와 영어 비디오 영상 한 개를 사용하여 제안한 방법을 검증하였다. 각 비디오 영상의 프레임은 640x480 해상도의 AVI 포맷이며, 데이터의 크기를 줄이기 위해서 1초에 8 프레임만을 샘플링하여 약 2,800개의 프레임으로 구성된 데이터를 사용하였다. 비디오 영상에서 하나의 동일한 자막은 최소한 8 프레임 이상, 즉 1초 이상 지속되었다. 실험은 동일 자막 판단의 정확성, 이미지의 향상 정도 및 문자 인식 결과에 초점을 두었으며, 처리 시간을 단축시키는 문제는 향후 연구에서 고려할 예정이다.

5.1 동일 자막 프레임 판단 실험

비디오 프레임을 분석하여 동일한 자막 프레임들을 판단하고 그룹화 시킨다. 표 3은 동일 자막 판단 과정을 수행하여 부여된 라벨의 개수, 즉 그룹의 개수를 보여준다. 각 비디오 데이터에 존재하는 서로 다른 자막의 개수, 동일 자막 판단 과정을 거쳐서 부여된 총 라벨의 개수와 소수 프레임 그룹을 제외시킨 최종적인 라벨의 개수를 각각 보여주고 있다. 최종 라벨을 갖는 프레임들만을 다중 결합한다. 한글 비디오1, 2는 ‘007 네버다이’,

표 3 동일 자막 판정에 의한 라벨링 결과

비디오 종류	총 자막 개수	부여된 라벨 개수	최종 라벨 개수(소수 라벨 제외)
한글비디오1	63	84	65
한글비디오2	72	72	66
영어비디오	128	331	115

표 4 라벨의 유형별 분포

비디오 종류	유형 (I)	유형 (II)	유형 (III)	유형 (IV)	유형 (V)
한글비디오1	44	9	6	3	3
한글비디오2	51	6	2	0	7
영어비디오	57	32	14	10	2

'아스테릭스'의 일부이고, 영어비디오는 '라이언 일병 구하기'의 일부이다.

표 4는 표 3의 실험 결과 중에서 최종 라벨 개수를 유형 별로 분류한 것이다. 유형 (I)은 동일한 자막 프레임들 모두에게 동일한 라벨이 부여된 경우로서 동일한 자막 프레임들로부터 정확하게 자막 영역이 추출된 경우이다. 유형 (II)는 동일한 자막 프레임들 가운데 일부만 동일한 라벨이 부여된 경우로서 중간의 일부 프레임들에서는 자막이 추출되지 않은 경우이다. 유형 (III)은 동일한 자막 프레임들에 두 개 이상의 서로 다른 라벨이 부여된 경우로서 추출된 자막 영역의 크기 등이 달라서 동일한 자막 프레임을 서로 다른 자막 프레임으로 판단한 경우이다. 유형 (I), (II), (III)은 모두 다중 결합을 수행하여 이미지를 개선시킬 수 있는 경우이다. 유형 (III)은 각각의 라벨을 갖는 프레임들을 다중 결합 한 후에 최종 자막 인식 결과를 활용하여 다시 동일한 그룹으로 수정하는 것이 가능한 경우이다.

유형 (IV)는 서로 다른 두 개의 자막 그룹에 동일한 라벨이 부여된 경우로서 다른 자막 프레임들을 동일한 자막 프레임으로 잘못 판단하여 발생한다. 유형 (V)는 동일한 라벨이 서로 다른 두 개의 자막 그룹의 일부 프레임들에 부여된 경우로서 역시 자막 영역이 올바르게 추출되지 못해서 발생한다. 유형 (III), (IV), (V)는 주로 자막 주위의 배경 변화가 심하여 예지가 조밀하게 추출되어 자막 영역을 부정확하게 추출하기 때문에 발생한다. 유형 (IV)와 (V)의 결과를 다중 결합하면 자막 이미지가 훼손되기 때문에 앞 절에서 언급한 재평가 과정을 수행하여 다시 분류한 후 결합한다.

5.2 이미지 향상 실험

5.2.1 다중 결합 결과

다중 결합 결과는 표 3의 유형 (I), (II), (III)에 해당되는 동일한 자막들이 결합된 결과와 유형 (IV), (V)에 해당되는 서로 다른 자막들이 결합된 경우로 분류할 수 있다. 동일한 자막들이 결합된 경우는 자막 영역은 보존되고 배경은 제거되는 효과를 볼 수 있다. 그러나 그림 9와 같이 서로 다른 자막들이 결합되면 오히려 자막이 훼손되는 결과를 만든다. 또한, 그림 10과 같이 배경의 변화가 있는 경우에는 다중 결합으로 배경 잡음을 대부분 제거할 수 있으나, 그림 11과 같이 배경의 변화가 거의 없는 경우에는 다중 결합으로 향상된 결과를 얻을 수 없었다.

유형 (IV)와 (V)의 다중 결합된 결과를 재평가하여 수정한 결과는 그림 12와 같다. 실험 데이터 중에서 한글비디오1과 영어비디오는 유형 (IV)와 (V)에 해당하는 자막들 모두가 그림 12와 같이 성공적으로 재분류되어 각각 다중 결합되었다.

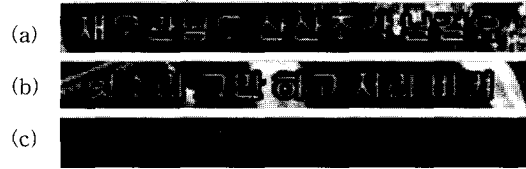


그림 9 서로 다른 자막들에 동일한 라벨이 부여된 경우의 다중 결합 : (a) 자막 1, (b) 자막 2, (c) 자막 1과 2에 해당되는 프레임들을 결합한 결과

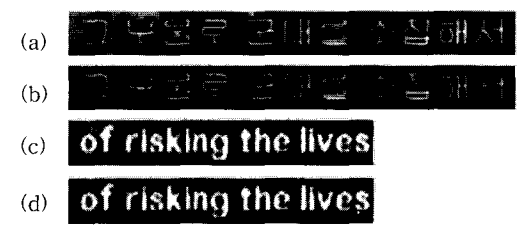


그림 10 배경의 변화가 있는 경우 : (a), (c) 결합 없이 추출한 경우; (b), (d) 결합한 후 추출한 경우

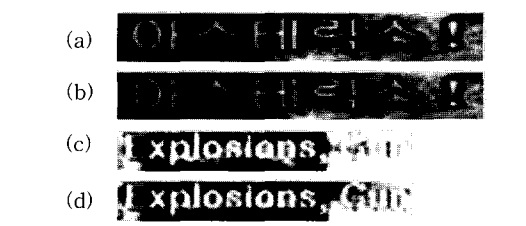


그림 11 배경의 변화가 없는 경우 : (a), (c) 결합 없이 추출한 경우; (b), (d) 결합한 후 추출한 경우

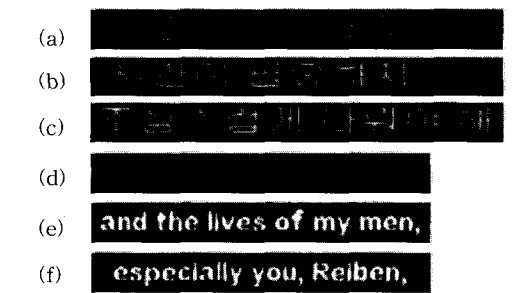


그림 12 재평가 수행에 의한 재분류 및 다중 결합 결과 : (a) 한글 자막 프레임의 재평가 및 재분류 전에 결합된 이미지; (b), (c) 재분류 후 결합된 결과; (d) 영어 자막 프레임의 재평가 및 재분류 전에 결합된 이미지; (e), (f) 재분류 후 결합된 결과

실험 영상 가운데 한글비디오2는 유형 (IV)와 (V)에 해당하는 자막들 중 단 한 개를 제외하고는 모두 성공적으로 재분류되었지만, 영어비디오는 동일한 자막 프레임들인데도 불구하고 배경의 급격한 변화로 인해서 두 그룹으로 재분류되는 결과가 한 번 발생하였다. 즉 재분류가 실패한 경우는 동일한 자막 프레임들이 서로 다른 자막으로 재분류되어 다중 결합되는 경우와 서로 다른 자막들이 재분류되지 못하는 경우로 나눌 수 있다. 첫 번째 경우는 그림 13과 같이 동일한 자막 프레임들의 배경이 급격하게 변하여 다중 결합한 후의 밝은 화소 값의 개수 변화가 크기 때문에 재분류가 발생한 경우이다. 프레임 번호 501에서 512까지 모두 동일한 자막 프레임이지만, 프레임 501과 프레임 502부터 512까지의 두 그룹으로 분리되고 각각 다중 결합되어 그림 13(a), (b)와 같이 재분류된 경우이다. 두 번째 경우는 그림 14와 같이 자막 영역의 크기와 문자들의 위치가 아주 비슷하여 다중 결합을 수행한 후에도 밝은 화소 값의 개수 변화가 거의 없기 때문에 재분류가 발생하지 않은 경우이다.

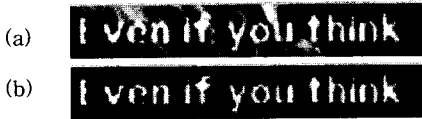


그림 13 동일한 자막들이 서로 다른 자막으로 판단되어 재분류된 후 결합된 결과

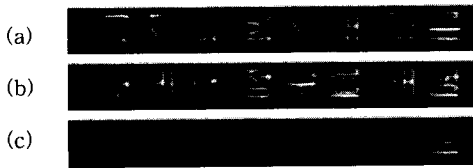


그림 14 서로 다른 자막들의 재분류가 실패한 경우

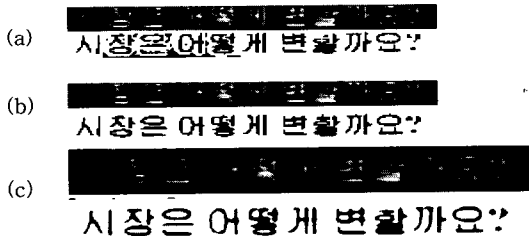


그림 15 (a) 원 이미지와 이진화 결과, (b) 다중 결합 이미지와 이진화 결과, (c) 해상도 증대 및 다중 결합한 이미지의 이진화 결과

5.2.2 해상도 증대 결과

그림 15는 해상도가 증대된 이미지와 그 이미지를 Niblack이 제안한 방법[19] 적용하여 이진화시킨 결과이다. 해상도 증대와 다중 결합을 수행하여 문자 외곽선 부분의 배경 잡영을 어느 정도 제거시킬 수 있음을 확인하였다.

5.2.3 히스토그램 평활화 결과

그림 16은 히스토그램 평활화가 적용된 결과 이미지를 보여준다. 그림 16(b)와 같이 이미지의 명도 값을 다시 분포시킨 이미지에서 문자 획에 해당되는 화소의 명도 값은 변화가 거의 없지만, 잡영을 포함하는 배경 화소의 명도 값은 낮아져서 결국 잡영의 일부가 제거되는 효과를 얻을 수 있다. 여기서 사용된 이진화 방법은 Niblack이 제안한 방법이다. 그림 16(b)에서 잡영이 추가로 제거된 결과를 볼 수 있지만, 복잡한 자소인 “ㅎ”과 “ㄹ”의 획들이 명확히 구분되지는 않는다. 이와 같이 자소의 획이 명확하게 구분되지 않으면 신뢰성 있는 인식 결과를 기대하기 어렵다.

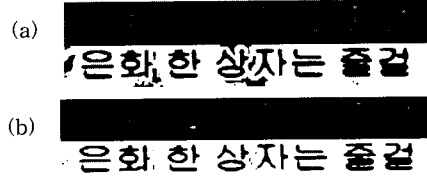


그림 16 (a) 다중 결합, 해상도 증대 후 이진화 결과, (b) 다중 결합, 해상도 증대 및 히스토그램 평활화 적용 후 이진화 결과

5.2.4 획 기반 이진화 결과

그림 17은 획 기반 이진화 결과를 보여준다. 그림 17(b)는 Niblack 이진화 방법을 적용한 결과 이미지로서 복잡한 자소인 “ㅎ”, “ㄹ”의 획들이 뭉쳐져 있는 것을 볼 수 있지만, 본 연구에서 사용한 Zhao 이진화 방법을 적용하면 그림 17(c)와 같이 이러한 획들이 보다 선명하게 추출됨을 확인할 수 있다.

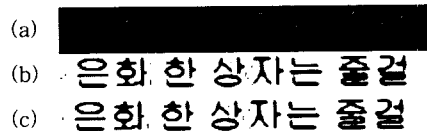


그림 17 (a) 다중 결합, 해상도 증대 결과, (b) (a)에 히스토그램 평활화와 Niblack 이진화 적용 결과, (c) (a)에 히스토그램 평활화와 Zhao 방법 적용 결과

5.2.5 스무딩 결과

스무딩에 사용된 모폴로지 연산은 Dilation과 Erosion 을 각각 한번씩 실시한 Opening 연산이다. 그림 18은 스무딩을 적용한 결과 이미지를 보여준다. 그림 18(b)에 있는 문자 획 주변에 나타나는 작은 잡영들이 제거된 것을 그림 18(c)에서 확인할 수 있다. 이 과정은 문자 인식 모듈에 별도의 이미지를 개선하는 전처리 과정이 포함되어 있다면 삭제할 수 있는 기능이다.

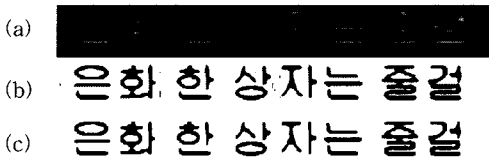


그림 18 (a) 다중 결합, 해상도 증대 이미지, (b) (a)에 히스토그램 평활화와 Zhao 이진화 방법 적용 결과, (c) (b)를 스무딩한 결과

5.2.6 이미지 향상 과정의 최종 결과 비교

그림 19는 다중 결합을 적용한 이미지와 적용하지 않은 이미지 각각에 대하여 해상도 증대, 히스토그램 평활화, 획 기반 이진화, 스무딩을 모두 적용한 최종 결과를 보여준다. 다중 결합을 먼저 수행한 결과 이미지에서 문자의 획들이 보다 선명하게 보존된 것을 확인할 수 있어서 좋은 인식 결과를 기대할 수 있다.

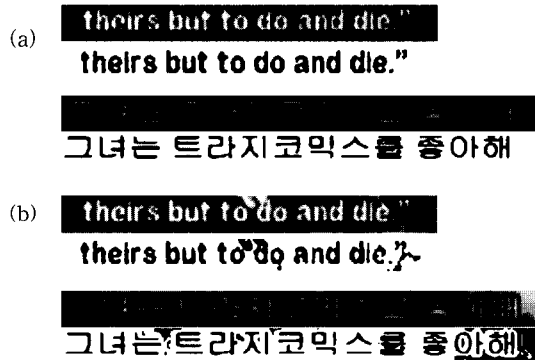


그림 19 (a) 다중 결합 후 단계적인 이미지 향상 방법 적용 결과, (b) 다중 결합을 수행하지 않고 단계적인 이미지 향상 방법 적용 결과

5.3 문자 인식 실험

4절에서 제안한 이미지 향상 방법들이 적용된 결과 이미지를 이진화 시킨 후 인식하였다. 인식은 추출된 자

막 이미지의 문자를 먼저 분리한 후 낱자 단위로 인식하였다. 영어 자막 인식은 상용 OCR 아르미4.0을 사용하였으며, 한글 자막 인식은 본 연구팀이 개발한 인식기를 사용하였다.

인식에 사용된 자막 이미지는 비디오 영상에서 추출한 영어 자막 100개와 한글 자막 221개이다. 영어 자막 100개에는 1,323자의 영어 낱자가 포함되어 있으며, 이미지 향상 방법을 단계적으로 적용한 인식 결과는 표 5와 같다. 예상한 것과 같이 이미지의 향상에 따라 인식 성능이 좋아진 것을 확인할 수 있다. 한글 자막 221개에는 1,496자의 한글 낱자가 포함되어 있으며, 이미지 향상 방법을 단계적으로 적용한 인식 결과는 표 6과 같다. 한글 자막 역시 이미지 향상에 따라 인식 성능이 크게 향상된 것을 확인할 수 있었다. 대부분의 오류는 복잡한 배경으로부터 발생되었으며, 이 경우 배경 화소는 아주 밝은 명도 값을 갖고 거의 움직이지 않아서 다중 결합으로 제거시키지 못했기 때문이었다. 이 실험에 사용된 인식 모듈은 모두 자막 인식에 최적화되어 있지 않기 때문에 추후 인식 모듈의 개선으로 더욱 정확한 인식 결과를 기대할 수 있다. 이 실험 결과는 단지 이미지 향상 과정의 필요성을 확인하기 위한 것이다.

표 5 영어 인식 결과(상용 OCR 사용)

이미지 향상 단계	인식률(%)
원 영상	27.4
다중 결합	58.3
해상도 증대	61.6
히스토그램 평활화	77.7
스무딩	81.2

표 6 한글 인식 결과(연구실 인식기 사용)

이미지 향상 단계	인식률(%)
원 영상	47.9
다중 결합	60.9
해상도 증대	73.2
히스토그램 평활화	80.9
스무딩	84.9

6. 결론 및 향후 과제

본 논문에서는 비디오 영상의 자막 이미지를 정확하게 인식하고 내용에 따른 인덱싱을 하기 위해서 동일한 내용의 자막을 갖는 프레임의 위치를 자동으로 찾는 방

법을 제안하였다. 또한, 이 결과에 따라 다중 결합과 해상도 증대, 히스토그램 평활화, 획 기반 이진화 및 스무딩 방법을 각각 제안하여 이미지의 수준을 향상시켰다. 영어 및 한글 비디오 자막에 제안된 방법을 적용하여 인식 성능이 향상된 것을 확인하였다. 이러한 연구 결과는 비디오 영상의 효율적인 색인에 활용될 수 있다.

향후 연구 과제로는 실시간 처리가 가능하도록 이미지 처리 시간을 단축시키는 것과 인식 모듈을 자막 폰트의 종류, 스타일 등에 맞추어 최적화시키는 것이다. 또한, 이미지 및 동영상에 자연스럽게 포함되어 있는 장면 텍스트 정보를 추출하는 방법을 개발하는 것이다.

참 고 문 헌

- [1] Y. Shong, H. Shang and A. K. Jain, "Automatic Caption Localization in Compressed Video", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 4, pp. 385-392, 2000.
- [2] Y. K. Lim, S. H. Choi and S. W. Lee, "Text Extraction in MPEG Compressed Video for Content-based Indexing", *Proceedings of 15th International Conference of Pattern Recognition*, Vol. 3, pp. 409-412, Spain, 2000.
- [3] 박상신, 김소명, 최영우, 정규식, "효율적인 비디오 자막 인식을 위한 영상 향상 방법", *제12회 영상처리 및 이해에 관한 워크샵 발표논문집*, pp. 342-346, 2000.
- [4] J.-C. Shim, C. Dorai and R. Bolle, "Automatic Text Extraction from Video for Content-Based Annotation and Retrieval", *Proceedings of 14th International Conference of Pattern Recognition*, Vol. 1, pp. 618-620, Brisbane, 1998.
- [5] H. Li and D. Doermann, "Automatic Identification of Text in Digital Video Key Frames", *Proceedings of 14th International Conference of Pattern Recognition*, Vol. 1, pp. 129-132, Brisbane, 1998.
- [6] Hae-Kwang Kim, "Efficient Automatic Text Location Method and Content-Based Indexing and Structuring of Video Database", *Journal of Visual Communication and Image Representation*, Vol. 7, No. 1-4, pp. 336-344, 1996.
- [7] 이미숙, 방건, 임영규, 홍영기, 김두식, 이성환, "내용 기반 색인 및 검색을 위한 실시간 뉴스 비디오 파서의 설계 및 구현", *한국정보과학회 가을 학술발표논문집*, Vol. 24, No. 1, pp. 365-268, 1997.
- [8] T. Sato, T. Kanade, E. K. Hughes and M. A. Smith, "Video OCR for Digital News Archives", *IEEE Workshop on Content-Based Access of Image and Video Databases*, pp. 52-60, India, 1998.
- [9] H. Li, O. Kia and D. Doermann, "Text Enhancement in Digital Video", *Proceedings of SPIE Conference on Document Recognition and Retrieval VI*, Vol. 3651, pp. 2-9, 1999.
- [10] V. Wu, R. Manmatha and E. M. Riseman, "Finding text in images", *2nd ACM International Conference on Digital Libraries*, Philadelphia, pp. 3-12, 1997.
- [11] 전병태, 배영래, 양영규, 오길복, "다단계 특징 추출에 의한 일반화된 자막 영역 추출 방법", *제12회 영상처리 및 이해에 관한 워크샵 발표논문집*, pp. 429-434, 2000.
- [12] 전병태, 정세윤, 이재연, 배영래, "뉴스 아이콘 자막 및 내용 자막 추출", *한국정보과학회 가을 학술발표논문집*, Vol. 24, No. 2, pp.127-130, 1997.
- [13] Pyeong-Kee Kim, "Automatic text location in complex color images using local color quantization," *Proceedings of the IEEE Region 10 Conference*, Vol. 1, pp. 625-632, 1999.
- [14] 이성환, "시각장애인을 위한 착용형 컴퓨터 시각 기술의 현황 및 전망," *제13회 영상처리 및 이해에 관한 워크샵 발표논문집*, pp. 3-8, 2001.
- [15] Yu Zhong, Kalle Karu and Anil K. Jain, "Locating text in complex color image," *Pattern Recognition*, Vol. 28, No. 10, pp. 1523-1535, 1995.
- [16] Anil K. Jain, Bin Yu, "Automatic text location in images and video frames," *Pattern Recognition*, Vol. 31, No. 12, pp. 2055-2076, 1998.
- [17] R. Gonzalez and R. Woods, *Digital Image Processing*, Addison Wesley, 1992.
- [18] M. Kamel and A. Zhao, "Extraction of Binary Character/Graphics Images from Grayscale Document Images", *Graphical Models and Image Processing*, Vol. 55, No. 3, pp. 203-217, 1993.
- [19] W. Niblack, *An Introduction to Image Processing*, Prentice Hall, 1986.



김 소 명

1999년 숙명여자대학교 전산학과 졸업(학사). 2001년 숙명여자대학교 컴퓨터과학과 졸업(석사). 2001년 2월 LG전자 디지털 TV 연구소 연구원. 관심분야는 영상처리, 컴퓨터 비전 등



곽 상 신

1995년 숭실대학교 전자공학과 졸업(학사). 2000년 숭실대학교 전자공학과 졸업(석사). 2000년 1월 Tech1system 연구원. 2001년 12월 OrCom 주임 연구원. 관심분야는 영상처리, 평판 디스플레이 등

최 영 우

정보과학회 논문지 : 소프트웨어 및 응용 제 29 권 제 1·2 호 참조



정 규 식

1979년 서울대학교 전자공학과 졸업(학사). 1981년 한국과학기술원 전산학과 졸업(석사). 1981년 2월 ~ 1984년 7월 금성사 중앙연구소 선임연구원. 1984년 9월 ~ 1990년 8월 University of Southern California 컴퓨터공학과 졸업(석사, 박사). 1993년 12월 ~ 1994년 3월 IBM Watson 연구소 방문연구원. 1998년 3월 ~ 1999년 2월 IBM Almaden 연구소 방문연구원. 1990년 9월 ~ 현재 숭실대학교 정보통신전자공학부 부교수. 관심분야는 인터넷컴퓨팅, 인공지능, 멀티미디어 등