論文
3-7-9

비디오 영상 정보 검색을 위한 문자 추출 및 인식 (Caption Detection and Recognition for Video Image Information Retrieval)

구 건 서*
(Gun-Seo Koo)

요 약

본 논문에서는 비디오에서 입력된 영상으로부터 내용기반 검색을 위해 자동으로 자막을 추출하여 특징 추출을 기반의 단층 연결 신경망 인식기(FE-MCBP)에 의해 자막 문자를 인식하여 영상 자막의 내용을 검출하는 방법을 제시하였다. 비디오에서 자막 추출은 먼저, 비디오에서 일정한 시간 간격으로 획득한 프레임 중에서 히스토그램 분석을 통하여 키 프레임을 찾는 과정을 수행하며, 그 다음에 각각의 키 프레임에 대하여 칼라 세그먼테이션 후 라인 검사 방법 통하여 자막 영역을 추출하도록 하였다. 마지막으로 추출된 자막영역에서 개별 문자를 분리하였다. 본 연구에서는 칼라 히스토그램을 분석 후 지역 최대값을 이용하여 세그먼테이션 후 라인 검사를 수행함으로써 처리 속도와 자막영역 검출의 정확도를 개선하였다. 비디오에서 자막 추출은 비디오 정보를 멀티미디어 데이터베이스화하는 초기 단계로 추출된 자막은 바로 문자 인식기의 입력이 된다. 또한 인식된 자막정보는 데이터베이스로 구축되며 내용기반 검색 기법에 의해 검색되도록 하였다.

ABSTRACT

In this paper, We propose an efficient automatic caption detection and location method, caption recognition using FE-MCBP(Feature Extraction based Multichained BackPropagation) neural network for content based retrieval of video. Frames are selected at fixed time interval from video and key frames are selected by gray scale histogram method. for each key frames, segmentation is performed and caption lines are detected using line scan method. lastly each characters are separated

This research improves speed and efficiency by color segmentation using local maximum analysis method before line scanning. Caption detection is a first stage of multimedia database organization and detected captions are used as input of text recognition system. Recognized captions can be searched by content based retrieval method.

1. 서론

최근 하드웨어와 통신 기술의 발달로 다양한 형태의 대용량 멀티미디어 정보의 저장, 검색, 처리가 가능하게 되었으며, 멀티미디어 데이터베이스가 디지털의 형태로 변환하면서 멀티미디어 데이터가 지닌 요구되었다. 이는 기존의 주석기반의 색인 방식의 문제점을 극복하기 위해 멀티미디어 데이터의 특성을 자동으로 추출하고 이를 기반으로 검색을 하는 내용기반 검색(Content-Based Technique)이 등장하게

* 정회원 : 숭의여자대학 인터넷 정보과 교수

논문접수 : 2002. 6. 20.
심사완료 : 2002. 7. 5.

※ 본 논문은 숭의여자대학 교내 연구비 지원에 의한 것임.

되었다[1][2].

기존 멀티미디어 정보의 효율적인 검색을 위하여 최근 여러 검색 기법들이 제안되고 있는데 이는 크게 다음 두 가지로 분류될 수 있다. 첫 번째 텍스트 기반 검색은 대상이 가지고 있는 개념이나 의미를 표현하는 것으로 빠른 검색 시간을 가지며 정확한 검색 효율을 얻는다. 그러나 텍스트 기반 검색 방법은 각 이미지들의 초기 텍스트 정보만을 검색에 사용하고 하나의 이미지 내용을 정확하게 텍스트로 표현하는데 있어 주관적 판단이 들어가기 쉽고 이미지의 모든 의미를 표현하는데 한계가 있다. 두 번째 이미지 내용 기반 검색은 색깔이나 모양과 같은 이미지의 특징을 표현하는 것으로 사용자의 질의 및 검색과정이 매우 편리하고, 인간의 시각적 인지과정에 의해 가시적으로 시스템과 상호작용 하면서 대화할 수 있다. 그러나 이미지 내용 기반 검색은 DB의 양이 증가할수록 검색 시간이 증가하며 유사도 측정에 의한 검색방법이므로 검색 효율이 떨어진다.

본 논문에서는 그 중에서 가장 많은 정보를 포함하고 있는 것은 텍스트를 이용하기 위해 비디오에서 자막을 추출하는 방법을 제시하였다. 따라서 본 논문에서는 비디오에서 입력된 영상으로부터 내용기반 검색을 위하여 자동으로 자막을 추출한 후 역전파(BackPropagation) 인식기에 의한 문자를 인식하여 영상 자막의 내용을 검출한 후, 이를 멀티미디어 데이터 베이스에 저장, 검색하는 방법을 제시하였다.

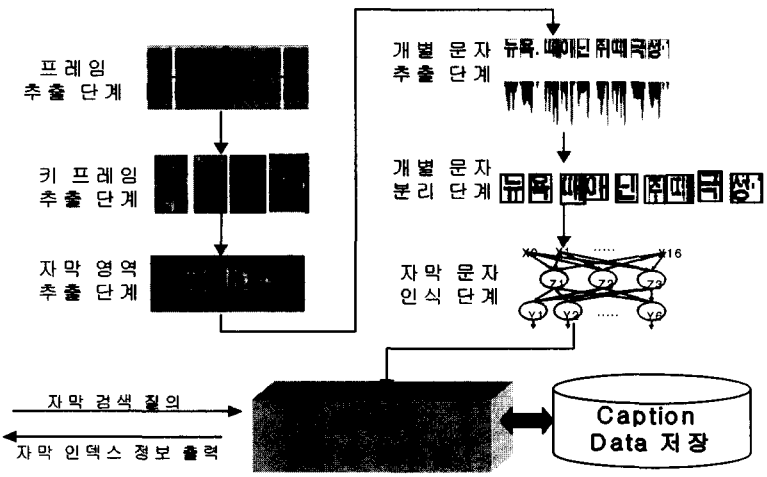
비디오에서 자막 추출은 먼저, 비디오에서 일정한 시간 간격으로 획득한 프레임 중에서 히스토그램 분석을 통하여 키 프레임을 찾는 과정을 수행하였다. 그 다음에 각각의 키 프레임에 대하여 칼라 세그먼테이션 후 라인 검사 방법 통하여 자막 영역을 추출하도록 하였다. 마지막으로 추출된 자막영역에서 개별문자를 분리하도록 하였다.

본 연구에서는 칼라 히스토그램을 분석 결과를 토대로 라인 검사를 수행함으로써 처리 속도와 자막 영역 검출의 정확도를 개선하고자 하였다. 비디오에서 자막 추출은 비디오 정보를 멀티미디어 데이터베이스화하는 초기 단계로 추출된 자막은 바로 문자 인식기의 입력부분이 된다. 인식된 자막정보는 데이터 베이스로 구축되며 내용기반 검색 기법에 의해 바로 검색될 수 있다.

2. 자막추출 시스템 설계 및 구현

2.1 시스템 구성 및 설계

자막인식을 위한 첫 단계로 뉴스를 녹화한 비디오에서 오버레이 보드를 통해 초당 1 프레임씩 획득하였다. 이때 획득된 프레임은 비압축 정지 영상이 되며, 다음 단계로 획득한 프레임 중에서 키 프레임



[그림 1] 자막 추출 시스템

[Fig. 1] Caption Extraction System

을 찾는다. 비디오에서 장면이 급전하는 경우에는 장면의 경계에 해당하는 영상간의 조명이나 배경 등이 크게 바뀌게 되며, 이때 장면의 경계에 해당하는 프레임을 컷이라고 하였다. 키 프레임 추출은 이러한 컷을 검출하기 위한 기술로써, 사용하는 영상의 특징에 따라서 화소 단위의 검출 방법, 부분 영역 단위의 검출 방법, 프레임 단위의 검출 방법으로 나누어진다. 키 프레임의 추출은 대용량의 비디오 데이터를 함축적으로 표현하기 위해 수행된다[2].

다음단계는 찾은 키 프레임 중에서 자막이 있을 경우 자막 영역을 추출하고 자막이 없는 경우에는 다음 키 프레임에 대하여 반복 처리를 하는 과정이다. 마지막 단계에서는 획득된 자막영역에서 개별문자 추출을 통해 자막을 문자 단위로 나눈 후 문자 인식기에 적용하여 자막을 인식하여 DB를 구축하였다.

[그림 1]은 본 논문에서 구현하는 자막 추출 시스템의 흐름도이다.

2.2 칼라 히스토그램을 이용한 키 프레임 검출

본 논문에서는 키 프레임을 검출하기 위하여 칼라 히스토그램을 사용하였다. 기존의 키 프레임 검출 방법 중에서 히스토그램을 사용한 검출 방법과 에지 영상을 사용한 검출 방법이 비교적 정확하게 컷을 검출하였다. 그러나 농도 히스토그램을 사용한 장면 전환 검출 방법은 영상의 농도값만을 특징으로 사용하기 때문에 서로 다른 장면의 프레임이라고 하더라도, 농도 히스토그램의 분포가 비슷하면 같은 장면으로 분류하게 되는 문제점이 있다.

에지 영상을 사용한 검출 방법은 영상의 에지를 구하기 위하여 Canny, Laplacian, Range, Sobel등을 오퍼레이터를 사용하는데 이것은 시간이 많이 소요되는 작업으로 에지 영상을 이용한 검출 방법이 여러 키 프레임 검출 방법 중에서 가장 시간이 오래 걸리게 하였다[2].

따라서 본 논문에서는 키 프레임 검출의 정확도와 속도를 고려하여 가장 적당한 방법으로서 칼라 히스토그램을 사용한 검출 방법을 적용하였다.

칼라 히스토그램을 사용한 장면 전환 검출 방법은 적용하는 칼라 모델에 따라서 RGB, YUV, HSI의 히스토그램을 이용하였다. 실험 결과 어떤 칼라 모델을 적용하던지 키 프레임 검출의 정확도는 크게 다르지 않았으나 YUV, HSI 칼라 모델을 적용하는 경우 영상을 대상 모델로 변환시키는데 많은 시간이 소요됐다. 키 프레임 추출 과정에서는 수많은 데이터 프레임들을 처리해야 하므로 처리 속도가 시스템 성능에 중요한 영향을 미친다. 따라서 처리 속도를 빠르게 하는 것이 중요한 문제가 된다. 특히 자막 프레임을 찾기 위하여 키 프레임을 추출하는 경우, 자막은 몇 초간 지속되므로 매우 정확한 장면전환 검출은 요구되지 않는다. 따라서 본 연구에서는 RGB 칼라 히스토그램을 이용한 검출 방법을 사용하였다. RGB 칼라모델은 red, green, blue의 세 개의 요소의 조합으로 칼라를 표현하는 방법으로 대부분의 컴퓨터 그래픽스 시스템들이 사용하는 칼라 모델로서 구현하기가 간편하다. RGB 칼라 히스토그림을 이용한 검출방법은 원영상에 대하여 바로 처리를 하므로 빠른 처리 속도를 가지고 있으며 또한 칼라 정보를 이용하므로 키 프레임 검출률도 우수하다.

RGB 칼라 히스토그램을 이용한 검출 방법은 다음과 같다.

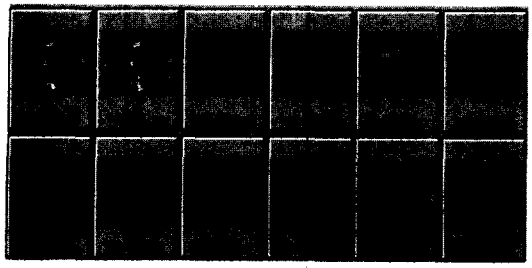
1단계 : 첫 번째 프레임에 대하여 RGB 각각의 히스토그램을 구한다.

2단계 : 두 번째 프레임에 대하여 RGB 각각의 히스토그램을 구한다.

3단계 : 프렘임 간의 RGB 각각의 히스토그램 차이의 누적 값을 구한다.

4단계 : RGB 히스토그램 중 하나라도 누적값이 임계치 이상이면 그 프레임을 컷으로 검출한다.

각 프레임은 비디오에서 초당 한 개씩 캡쳐된 것으로 시간순으로 순차적으로 변화하고 있음을 알 수 있다. 추출된 프레임들은 각각 영상이 비슷하기 때문에 프레임 수에 비하여 많은 정보를 표현하고 있지 못하다는 것을 알 수 있다. [그림 2]에서 추출한 프레임가운데에서 [그림 3]은 흑백 히스토그램을 이용하여 추출한 키 프레임들을 이며, 자막 프레임은 키 프레임의 부분 집합으로 정보를 함축적으로 표현하고 있다.



[그림 2] 비디오에서 추출된 프레임

[Fig. 2] Extracted Frames from Video Images

[그림 3] 추출한 키 프레임

[Fig. 3] Extracted Key Frames

2.3 자막 영역 검출

자막 영역 검출은 추출한 키 프레임을 대상으로 그 프레임에 자막이 있을 경우 자막 영역만을 추출하는 것을 그 목적으로 한다. 자막 영역 검출은 위와 같은 자막 영역의 특징을 이용하여 다음과 같은 방법으로 수행할 수 있다.

입력된 키 프레임에 대하여 칼라 세그먼테이션 수행 후 영역별 라인검사를 수행하여 자막 후보영역을 결정하였다. 후보 영역이 자막영역으로 결정되면 자막영역을 검출하고 다음 단계인 자막추출을 수행하고 그렇지 않은 경우에는 입력된 키 프레임이 자막 프레임이 아니라는 메시지를 내보내고 위의 과정은 반복하였다.

2.3.1 칼라 세그먼테이션

칼라 세그먼테이션은 일반적인 컴퓨터 비젼 및 패턴 인식분야에서 매우 중요하고도 비용이 많이 소요되는 처리단계이다. 수백 가지의 다양한 기법의 칼라 세그먼테이션 방법이 이진화, 클러스터링 또는 통계적 방법에 의한 에지 검출, 퍼지 이론과 신경망에 기초하여 연구되었다. 그러나 모든 어플리케이션에 적당한 세그먼테이션 방법은 아직 없는 상황이다.

비디오 이미지에 있어서 각 영상들은 다른 색을 가진 몇 개의 물체들로 이루어져 있으므로 정확한 자막영역 검출을 위해서는 칼라 세그먼테이션이 필요하다. 칼라 모델은 칼라와 그것들의 관계를 표현하는 방법으로 다른 영상처리 시스템은 다른 칼라모델을 사용하였다. 칼라 모델의 결정은 칼라 세그먼테이션의 결과에 큰 영향을 미친다. 주요 칼라 모델로는 RGB 모델, CMY/CMYK 모델, HSI 모델이 있다. RGB 칼라모델은 Red, Green, Blue의 세 개의 요소의 조합으로 칼라를 표현하는 방법으로 대부분의 컴퓨터 그래픽스 시스템들이 사용하는 칼라 모델로서 구현하기가 간편하다. Red, Green, Blue의 칼라 요소는 서로 심하게 간섭을 하므로 몇몇 영상처리를 어렵게 하는 단점이 있다.

CMY 칼라모델은 청록색(Cyan), 자홍색(Magenta), 노랑색(Yellow)의 세 요소로 이루어져 있다. CMY 칼라모델은 칼라 영상의 인쇄물들을 나타낼때 쓰이는 방법이다. CMY 모델에 검은색(blacK)요소를 더한 것이 CMYK 모델로서 인쇄의 품질을 높이기 위해 사용된다. HSI 칼라모델은 H는 색조(Hue), S는 채도(Saturation)를 I는 명암(Intensity) 을 이용하여 칼라를 표현하였다. HSI 모델에서는 red, blue, green 요소의 확률을 고려할 필요가 없으며 단지 원하는 색이 있으면 Hue를 적용하고, 진한 빨강을 핑크색으로 바꾸고 싶다면 Saturation을, 밝기를 조절하고 싶다면 Intensity를 바꾸면 된다. RGB영상을 HSI로 변화시키는 방법은 다음과 같다.

$$I=\frac{1}{3}(R+G+B) \tag{1}$$

$$S=1-\frac{3}{(R+G+B)}[\min(R+G+B)]$$

$$H=\cos^{-1}\left[\frac{\frac{1}{2}[(R-G)+(R-B)]}{\sqrt{(R-G)^2+(R-(G-B)\ B}}\right]$$

CMY/CMYK 모델은 영상 인식을 위한 처리 모델로는 적합하지가 않으며 HSI 모델은 인간의 칼라 인식 개념에 부합하는 장점이 있으나 RGB 영상을 HSI 영상으로 변환시키는 데에는 복잡한 계산이 필요하므로 많은 시간이 소요된다. 또한 거리 계산에

있어서도 RGB 모델보다 적합하지 않다. 따라서 본 논문에서는 처리 속도를 빠르게 하기 위하여 RGB 모델을 사용하였으며 거리 계산은 유클리드 거리를 이용하였다.

칼라 세그먼테이션 수행방법은 다음과 같다.

1단계 : R, G, B 별로 각각의 칼라 히스토그램을 구한다.

2단계 : 영역별로 히스토그램의 최대치를 구하였다. 영역 분할시에 밝은 부분에 가중치를 주어서 나눈다.

3단계 : 각 영역별로 local maximum을 구한다.

4단계 : 영상의 픽셀값을 각 local max와 거리의 차가 가장 적은 쪽으로 근사시킨다. 여기서 거리는 유클리드 거리를 이용하여 구하였다. 즉, local max(R1, G1, B1)와 얻어진 pixel(R2, G2, B2)에 대하여 두 픽셀간의 거리 d는 식(2)와 같이 구해진다.

$$d=(R1-R2)^2+(G1-G2)^2+(B1-B2)^2 \quad (2)$$

칼라 세그먼테이션 수행시 영역을 몇 개로 나누는가가 시스템의 성능에 중요한 요소로 작용하였다. 위와 같은 방법으로 칼라 세그먼테이션을 수행하면 local max의 개수가 언제나 일정하므로 영상의 특성에 구애받지 않고 빠른 칼라 세그먼테이션을 수행할 수 있는 장점이 있다.

[그림 4]는 추출된 키 프레임의 칼라 히스토그램을 보여준다. [그림 4](b)는 [그림 4](a)의 칼라 세그먼테이션의 결과이다. 위의 그림과 비교하여 히스토그램 분포도가 변화했음을 알 수 있다.

2.3.2 영역별 라인 검사

1단계 : 처리속도의 개선을 위하여 칼라 세그먼테이션 된 영상을 그레이 스케일 영상으로 변환한다.

2단계 : 변환된 영상을 수평 방향으로 한 라인씩 영역별로 히스토그램을 구하였다. 영역을 나누는 이유는 라인 검사의 특성상 문자의 개수가 적을 시에는 문자 영역을 검출하지 못하기 때문에 사람의 이름처럼 가운데 있는 문자영역도 검출할 수 있도록 하기 위함이다.

3단계 : 히스토그램의 분포도를 조사하여 각 영역별 변화의 누적값을 기록한다.

4단계 : 변화가 발생한 X 좌표의 위치를 기록한다.

5단계 : 히스토그램의 변화값의 프로파일을 작성한다.

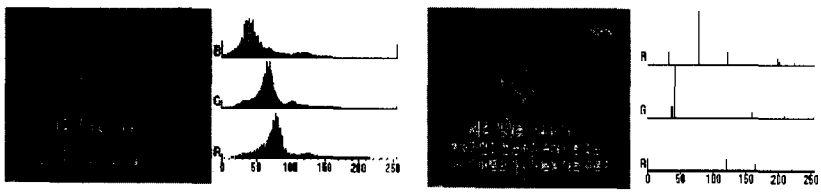
6단계 : 밝은 픽셀의 개수와 평균 변화량을 계산한다.

2.3.3 후보 라인결정

1단계 : 한 영역이라도 변화의 누적값이 임계치를 만족하면 그 라인을 자막의 후보라인으로 선택하였다. 자막 후보라인을 구하는 식은 다음과 같다.

$$H_{MIN} < line\ variance < H_{MAX} \quad (3)$$

영상이 세그먼테이션 된 영상이므로 변화도가 너무 크거나 작으면 자막 영역이



(a) 칼라 히스토그램 (b) 칼라 세그먼테이션

(a) Color Histogram (b) Color Segmentation

[그림 4] 히스토그램의 분포도

[Fig. 4] Distribution Chart of Histogram

아니므로 위와 같은 식이 유도된다. 위 식에서 두 개의 임계값 HMIN과 HMAX 는 경험에 의해 결정한다.

2 단계 : 히스토그램의 프로파일이 평균변화량보다 높은 라인을 후보라인으로 결정한다.

위의 두 조건중 하나라도 만족하면 후보라인으로 결정한다.

2.3.4 후보 영역결정

1단계 : 후보 라인 중 히스토그램의 변화가 갑자기 상승하는 스파크 현상이 없으며 밝은 부분의 값이 상대적으로 높은 영역을 찾는다.

2단계 : 각 후보라인별로 위치를 구한다.

3단계 : 구해진 위치정보를 가지고 영역을 조건에 따라 병합한다. 이때 영역의 면적과 후보 영역들 사이의 거리를 고려해서 임계치를 만족하는가를 검사한다.

4단계 : 마지막으로 면적을 조사하여 조건을 만족하는 영역을 자막의 후보영역으로 결정한다.

[그림 5](a)은 라인 스캔하여 조건에 맞는 라인에 선을 그어서 문자 후보 라인을 표시한 그림이다. 문자 영역에 선이 밀집되어 있지만 비 문자 영역에도 선이 그어져 있는 것을 볼 수 있다. [그림 5](b)는 후보영역 결정의 결과를 보여준다. 문자 영역이 아니지만 라인 검출이 된 부분이 제거되었음을 알 수 있다.

2.3.5 자막영역 결정

1 단계 : 각 후보영역별로 수직 방향으로 스캔한다.

2 단계 : 한 영역이라도 변화의 누적값이 임계치를 만족하면 그 부분을 자막 영역으로 선택한다. 자막영역을 결정하는 식은 식(3)과 같다.

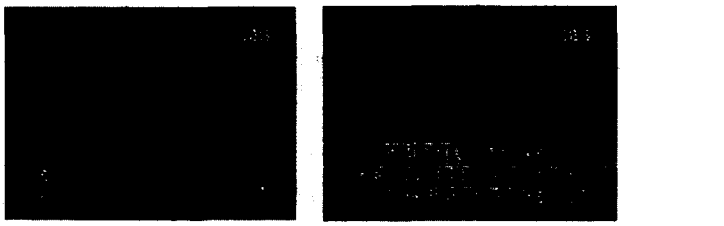
영상이 세그먼테이션 된 영상이므로 변화도가 너무 크거나 작으면 자막 영역이 아니므로 위와 같은 식이 유도된다. 위 식에서 두 개의 임계값 V_{MIN}과 V_{MAX}는 경험에 의해 결정한다.

3단계 : 마지막으로 후보영역의 프로파일이 주변의 프로파일보다 상대적으로 높은 경우 자막영역으로 결정한다.

위와 같은 처리 방법은 전처리로써 칼라 세그먼테이션을 수행하여 영상을 영역별로 분할한 결과를 가지고 라인스캔하므로 자막영역 검출률을 향상시키고 오검출의 횟수를 감소시킬 수 있다.

2.4 자막 추출

위에서 추출된 자막 영역은 바로 문자 인식기에서 인식할 수 없다. 따라서 자막의 추출은 문자 인식기에서 인식할 수 있도록 획득된 흑백의 자막 영상을 이진화 하여 문자 단위로 분할하는 것을 그 목적으로 한다. 자막 추출 방법은 다음과 같다.



(a) 영역별 라인 스캔 (b) 후보영역 결정
(a) Local line scan (b) Decide Candidate Domain

[그림 5] 자막 후보 영역 검사
[Fig. 5] Caption Candidate Region Check

1단계 : 일반적으로 자막은 흰색계열의 밝은 색이기 때문에 영상을 반전시킴으로써 글자부분을 강조시킬 수 있다. [그림 6](a)는 추출된 자막영역이며, [그림 6](b)는 영상을 반전시킨 이미지이다. 글자 부분이 검은색으로 바뀌었음을 볼 수 있다.

2단계 : 반전된 영상을 히스토그램의 분포도에 따라서 P-tile 또는 Mode법을 사용하여 이진화 한다. [그림 6](c)은 이진화시킨 영상을 보여준다.

3단계 : 구해진 영상은 잡음이 있고 선의 굵기가 고르지 못 하므로 영상개선이 필요하다. 따라서 고립점을 제거하기 위하여 <표 1>과 같이 3 * 3 마스크 연산자를 이용하여 영상을 개선시킨다.

<표 1> 3 * 3 마스크

<Table> 3 * 3 Mask

a1	b1	c1
a2	b2	c2
a3	b3	c3

영상에서 수평과 수직 방향에 대하여 다음과 같이 누적값을 구한다.

T_h = a1 + b1 + c1 + a3 + b3 + c3, Tv = a1 + a2 + a3 + c1 + c2 + c3

if $T_h \leq 255*3$ or $T_v \leq 255*3$ then Pixcolor= 0 else Pixcolor = 255

흰색은 255이고 검은색은 0이므로 다음과 같이 고립점 여부를 판단한다. [그림 6](e)는 고립점을 제거하여 개선된 영상을 보여준다.

4단계 : 개선된 영상을 가로방향과 세로방향으로 라인 단위의 히스토그램 검사를 수행한다.

5단계 : 검색 결과에 임계치를 적용하여 문자 후보를 찾는다.

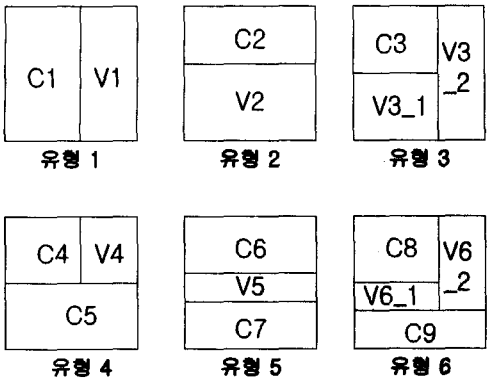
6단계 : 후보 문자에 휴리스틱을 적용하여 문자별로 분할한다. [그림 6](g)는 투영을 통하여 개별문자를 분할한 영상을 보여준다.

[그림 6] 자막 문자 추출 및 인식

[Fig. 6] Extract and Recognition for Caption Character

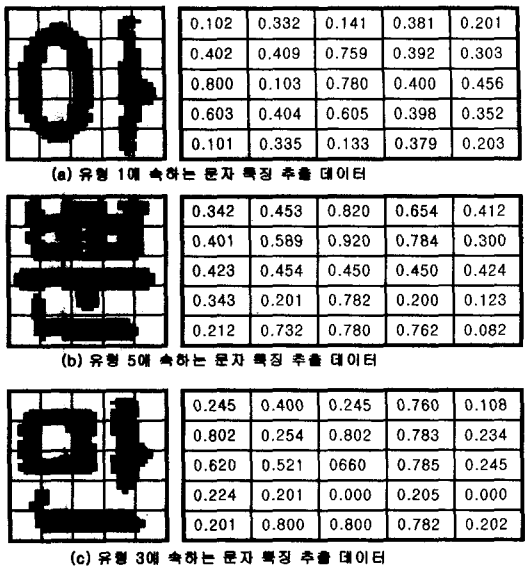
2.5 문자 인식

한글은 문자수가 많고 초성, 중성, 종성이 2차원으로 조합되어 하나의 문자를 이룬다. 초성이 19개, 중성이 21개, 그리고 종성이 27개로서 이들의 조합으로 가능한 모든 문자의 수는 11,172자이고 일반적으로 많이 사용되는 문자만도 완성형 코드 체계에서 제공되는 2,350자나 된다. 한글은 자음과 모음이 조합되는 종류에 따라 [그림 7]과 같이 6개의 유형으로 나눌 수 있다[5][13].



[그림 7] 한글의 여섯가지 유형

[Fig. 7] 6 type's of Hangul



0.102	0.332	0.141	0.381	0.201
0.402	0.409	0.759	0.392	0.303
0.800	0.103	0.780	0.400	0.456
0.603	0.404	0.605	0.398	0.352
0.101	0.335	0.133	0.379	0.203

(a) 유형 1에 속하는 문자 특징 추출 데이터

0.342	0.453	0.820	0.654	0.412
0.401	0.589	0.920	0.784	0.300
0.423	0.454	0.450	0.450	0.424
0.343	0.201	0.782	0.200	0.123
0.212	0.732	0.780	0.762	0.082

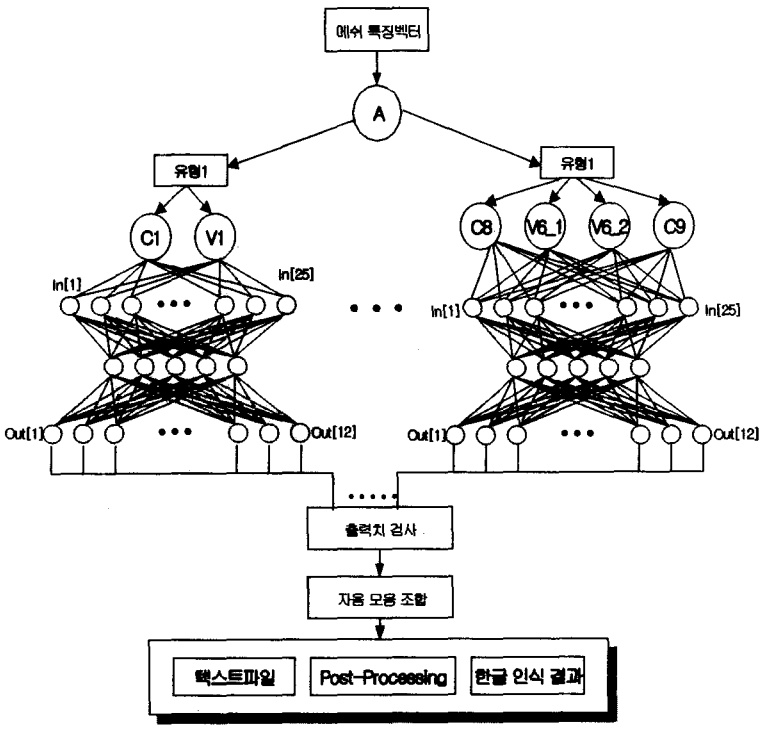
(b) 유형 5에 속하는 문자 특징 추출 데이터

0.245	0.400	0.245	0.760	0.108
0.802	0.254	0.802	0.783	0.234
0.620	0.521	0660	0.785	0.245
0.224	0.201	0.000	0.205	0.000
0.201	0.800	0.800	0.782	0.202

(c) 유형 3에 속하는 문자 특징 추출 데이터

[그림 8] 문자 특징 추출

[Fig. 8] Character Feature Extraction



[그림 9] 문자 인식 시스템

[Fig. 9] Character Recognition System

본 논문에서 문자 인식 처리는 비디오 영상 처리 과정을 통해 입력된 자막 문자 영상 부분이 개별 문자화하여 한 문자씩 문자 인식 단계에서는 이를 인식하여 그 결과를 출력하게 된다. [그림 9]와 같이 자막 문자를 인식 처리하는 시스템은 역전파 신경망을 기반으로 한 FE-MCBP(Feature Extraction based Multichained BackPropagation) 시스템을 응용하여 자막 인식을 위한 인식기로 적용하고자 한다[1]. FE-MCBP는 순차 조합의 장점인 신속성과 병렬 조합의 장점인 정확성을 결합한 하이브리드(hybrid) 조합 방법에 의해 처리되어지는 효율적 인식기로 인정되었다[1].

이와 같은 인식기의 특징을 이용하여 문자 분리 단계에서 구해진 메쉬의 특징벡터를 신경망의 입력으로 사용하여 문자를 인식하였다. 신경망은 유형 분류를 위한 1개의 신경망과 문자의 형태에 따른 6가지 신경망에 자음, 모음 인식을 위한 신경망으로 구성되어 있다. 신경망의 구조는 은닉층이 3 layer 구조를 갖는다. 신경망 A의 입력값으로는 메쉬의 64개의 특징벡터가 사용된다.

<표 2> 학습 및 인식 제어 파라미터

<Table 2> Control Parameter for Learning and Recognition

학습 제어 파라미터	인식 제어 파라미터

//file:korchartrn.rsp
// input units
-i=25
// output units
-o=12
// hidden layer
-hh=2
// 1st hidden layer
-h=2
// 2st hidden layer
-h=25
// 3nd hidden layer
-h=12
// recognition file
-ftrain=numtrn.trn
// dump file
-fdump=c25.dmp
//samplerecognition patterns
-samp=353
// report training status every 10 cycles
-r=10000
// time the training process
-t
// initialize the random weight(-0.5 to +0.5)
-w+=0.5
-w-=-0.5
// mean square per unit err = 0.001
-err = 0.001
// tolerance = 0.01 default = 0.001
-torerr=0.00001
	//file:korcharrgn-1.rsp
	// input units
	-i=25
	
	// output units
	-o=10
	
	// hidden layer
	-hh=3
	
	// 1st hidden layer
	-h=6
	
	// 2nd hidden layer
	-h=12
	
	// recognition file
	-frecog=n25rgn.rgn
	
	// dump file
	-fdump=n25trn..dmp
	
	// sample recognition patterns
	-samp=353
	
	// output file
	-fout=number.out

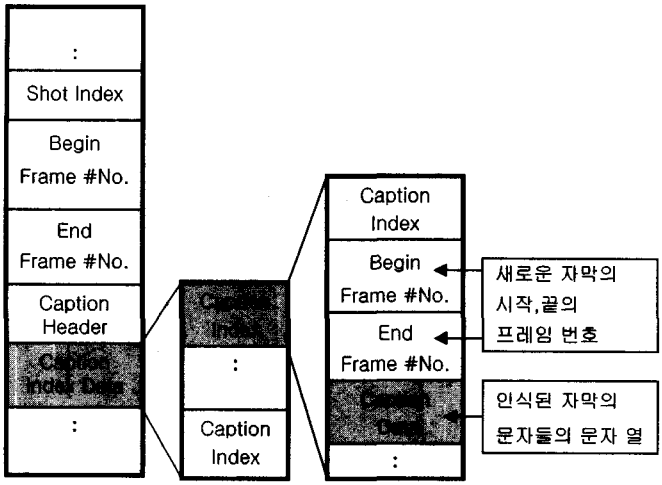
메쉬 특징벡터의 값을 입력으로 받아들여 일차적으로 여섯가지 유형 중 속하는 유형을 판단하고 결정된 유형에 따라 인식 문자의 범위를 결정짓는다. 이러한 방법을 적용하는 이유는 오인식률을 줄이고, 후처리 작업 시 자막 문자열 오 인식 문자열의 보정 처리할 수 있다는 장점을 갖고 있다. 기존 신경망 문자 인식 방법과 다른 방법은 개별 문자열을 5× 5 mask를 이용하여 2진 영상 정보의 평균값을 추출하여 한 결과가 문자인식을 위한 특징 값으로 사용된다. 문자 인식 시스템 구성은 [그림 9]과 같으며, <표 2>는 신경망 학습 및 인식 제어 파라미터 예문이다.

2.6 자막 데이터 저장 및 검색

비디오의 자막을 추출하여 인식한 결과는 장면 전환에 따른 프레임이 시작하는 시작 정보과 끝 정보 그리고 자막 문자 정보이다. 이러한 자막에 관한 정보는 [그림 10]과 같다. 저장된 인덱스 정보는 비디오 검색 시스템에서 내용 및 자막에 대한 질의를 이용하여 검색할 수 있으며, 검색 결과에 대한 정보를 제공하였다. 비디오 인덱싱 결과를 샷과 자막의 인덱스 구조로 분할하여 저장하였다.

비디오 인덱싱 결과로 검출된 장면 전환을 저장하기 위한 구조이다. 샷 인덱스 구조에서 시작과 끝 프레임 번호는 하나의 샷이 시작하고 끝나는 프레임 번호이다. 자막 헤더는 샷에 포함된 자막 종류의 개수를 의미하며, 자막 인덱스 데이터에 포함되어 있는 자막 인덱스는 자막 인덱스 구조와 매칭을 위한 위치 정보를 의미한다.

자막 인덱스 구조는 자막 프레임 검출, 자막 문자 인식 과정을 통해 얻어진 자막 정보를 저장하기 위한 것이다. 시작과 끝 프레임 번호의 의미는 자막이 나타나고 사라지거나 다른 자막으로 바뀌는 프레임 번호이다. 자막 문제 데이터는 인식된 자막 문자들의 문자열이다.



[그림 10] 자막 정보 저장 구조

[Fig. 10] Caption Information Storage Structure

3. 실험 및 평가

3.1 실험환경 및 방법

본 논문의 실험을 위하여 영상 입력 장비로는 LG LV-S22 7 head VTR과 두인전자의 윈도우 비전을 사용하였다. 운영체제로는 Windows 98를 사용하였고 컴파일러로는 Visual C++ 6.0을 사용하여 팬티엄 4/ Main Memory 1.7 GHz 칩이 장착된 PC에서 수행하였다.

데이터 비디오로는 SBS 8시 뉴스, KBS 8시 뉴스, KBS 9시 뉴스, MBC 9시 뉴스를 이용하였으며, 입력 영상으로는 640 * 480 크기의 칼라영상을 이용하였다.

실험 환경은 VTR로 재생한 화면을 오버레이 카드를 이용하여 초당 1 프레임씩 획득한 640 * 480 크기의 칼라 영상을 소프트웨어적인 영상처리 기법을 사용하여 처리하도록 하였다.

3.2 평가

본 논문의 평가를 위한 착안점은 크게 키 프레임 검출 부분과 자막영역 검출부분, 개별문자 추출 부분으로 나눌 수 있다. 그리고 본 논문에서 제안한 방법이 뉴스 비디오 영상뿐만 아니라 그림이 있는 다양한 배경의 문서에서 문자행을 추출하는데 적합한가를 알아보기 위해 다양한 배경의 문서 이미지에 대하여도 실험하였다.

키 프레임 검출은 많은 양의 데이터를 처리해야 하므로 정확한 검출뿐만 아니라 처리 속도도 중요한 문제이다. 따라서 키 프레임 추출은 처리속도와 정확도를 비교함으로써 그 성능을 비교할 수 있다. <표 3>은 대표적인 키 프레임 검출 방법인 흑백 히스토그램을 사용한 검출 방법과 본 논문에서 적용한 칼라 히스토그램을 이용한 검출 방법의 검출률과 속도를 비교하여 보여준다.

Nt는 데이터에 존재하는 실제 키 프레임의 개수, Nd는 정검출된 키 프레임 개수, Ne는 오검출된 키 프레임의 개수이고, c와 g는 각각 칼라 세그먼테이션을 사용한 방법과 흑백 히스토그램을 사용한 방법을 나타낸다. 검출률 R은 전체 프레임의 개수에서 정검출된 프레임의 비율로서 다음과 같이 계산하였다.

$$R = \frac{N}{Nt} * 100$$

정확도 C는 전체 추출된 프레임 중에서 정검출된 프레임의 개수로서 다음과 같이 계산된다.

$$C = \frac{Nd}{Ne + Nd} * 100$$

<표 3> 키 프레임 검출 결과

<Table 3> Result of Key Frame Extract

인덱스 / 뉴스	Nt	정검출		오검출		검출률		정확도	
		Ndc	Ndg	Nec	Neg	Rc	Rg	Cc	Cg
MBC 9시 뉴스	110	110	105	9	9	100	95.4	92.4	92.1
KBS 9시 뉴스	102	102	99	8	7	100	97.0	92.7	85.3
KBS 8시 뉴스	112	110	105	5	5	98.2	93.7	95.6	95.4
SBS 8시 뉴스	118	118	116	4	3	100	98.3	96.7	97.4

<표 4> 자막영역의 검출결과

<Table 4 > Result of Caption Region Extract

인덱스 / 뉴스	Nt	정검출		오검출		미검출		검출률	
		Ndl	Ndc	Nel	Nec	Nnl	Nnc	Rl	Rc
MBC 9시 뉴스	152	140	120	4	7	12	32	92.1	78.9
KBS 9시 뉴스	103	99	94	9	15	4	9	96.1	91.2
KBS 8시 뉴스	113	109	100	8	20	4	13	96.4	88.4
SBS 8시 뉴스	98	97	95	7	11	1	3	98.9	96.9

<표 5> 개별문자 추출 결과

<Table 5> Result of Individual Character Extract

인덱스 / 뉴스	Nt	정분류	오분류	검출률
MBC 9시 뉴스	502	490	42	97.6
KBS 9시 뉴스	510	491	49	96.2
KBS 8시 뉴스	535	503	50	94.0
SBS 8시 뉴스	587	545	62	92.8

<표 3>을 보면 두 방법모두 키 프레임 검출에 적당한 것으로 나타났으나 칼라 히스토그램을 사용한 방법이 흑백 히스토그램을 사용한 방법보다 검출률과 정확도가 더 우수한 것을 알 수 있다. 특히 칼라 히스토그램을 이용하는 경우 미검출되는 키 프레임이 거의 없다는 장점이 있다.

자막영역의 검출은 자막 추출 시스템에서 가장 중요한 부분으로 기존의 방법과 비교해서 정검출률을 높이면서 오검출률 줄이는 측면에서 비교되어야 한다. 자막영역은 배경처리가 된 경우와 않된 경우 검출률에서 차이가 나기 때문에 각각 분리하여 비교 검토되어야 한다.

<표 4>는 본 논문에서 제안한 검출방법과 문자 부분이 고주파라는 사실을 이용한 검출 방법으로 배경처리가 된 자막 영역을 검출했을 때의 결과를 보여준다. Nt는 데이터에 존재하는 실제 자막영역의

개수이고, Nd는 정검출된 자막영역의 개수이고, Ne는 오검출된 자막영역의 개수, Nn은 자막영역이지만 검출하지 못하는 미검출의 개수를 나타낸다. 그리고 l은 문자 부분이 고주파라는 사실을 이용한 검출방법을 나타내며, c는 본 논문에서 제안한 검출방법을 나타낸다. 검출률 R은 전체 자막영역의 개수에서 정검출된 자막영역의 비율을 백분율로 나타낸다.

<표 4>를 살펴보면 두 방법모두 검출률에서 우수한 결과를 나타내고 있다. 그러나 본 논문에서 제안한 방법이 문자 부분이 고주파라는 사실을 이용한 방법보다 오검출과 미검출에서 우수하다는 것을 알 수 있다.

<표 5>는 개별문자 추출의 실험결과를 보여준다. Nt는 데이터에 존재하는 문자의 개수이고, 정분류는 정확히 문자부분을 추출한 경우이고 오분류는 문자가 아니지만 문자로 분리하거나 문자부분을 정확히 분리하지 못하는 경우를 나타낸다. 검출률은 전체 문자의 개수에서 정분류된 문자의 개수를 백분율로 나타낸다.

3.3 기대효과

본 논문은 뉴스 비디오에서 자막을 추출하는 방법을 다루었다. 자막영역을 추출하기 위하여 칼라 세그먼테이션 후 라인 검사하는 방법을 사용하였는데 이러한 기법은 일반 비디오 영상뿐만 아니라 그림이 있는 다양한 배경의 문서에서 문자 행을 추출하는데 효과적으로 적용할 수 있을 것으로 기대된다. 뉴스 비디오에서 자막을 추출함으로써 얻을 수 있는 기대효과는 다음과 같다.

첫째, 기존의 순차적인 검색 방법을 탈피하여 내용기반 검색 기법에 의하여 빠른 검색을 할 수 있다. 둘째, 사람이 키보드를 사용하여 정보를 입력하는 데 드는 비용을 절약할 수 있다. 셋째, 검색된 정보는 일반 컴퓨터 텍스트이므로 정보의 재활용이 가능하다.

4. 결론

본 논문에서는 내용기반 검색을 위하여 뉴스 비디오에서 자막을 추출하는 방법을 제시한다. 키 프레임 추출시 RGB 칼라 히스토그램을 이용함으로써 높은 검출률뿐 아니라 속도를 빠르게 한다.

자막 영역 추출은 칼라 세그먼테이션 후 라인별로 히스토그램 변화를 분석한 프로파일을 이용함으로써 행한다. 전처리로서 칼라 세그먼테이션을 수행함으로써 오검출을 줄일 수 있었으며 또한 배경처리가 된 자막영역뿐만 아니라 배경처리가 않된 자막영역도 정확히 검출할 수 있었다. 칼라 세그먼테이션 수행시 지역 최대값을 이용하여 세그먼테이션을 수행함으로써 기존의 칼라 세그먼테이션 방법보다 처리 속도를 개선하였다.

추출된 자막영역은 이진화 시킨 후 수직 방향으로 투영하여 투영값에 대한 히스토그램을 구한 다음, 히스토그램값이 임계치 이하인 곳에서 문자를 분리하였다. 자막영역의 이치화는 히스토그램의 분포에 따라 P-tile법과 Mode법을 병행하여 처리함으로써 보다 좋은 결과를 얻을 수 있었다.

비디오에서 자막 추출은 비디오 정보를 멀티미디어 데이터베이스화하는 초기 단계로 추출된 자막은 바로 문자 인식기의 입력부분이 된다. 인식된 자막정보는 데이터베이스로 구축되며 내용기반 검색 기법에 의해 바로 검색될 수 있다.

본 연구에서는 비압축 비디오를 대상으로 자막을 추출하는 방법을 제시하였다. 본 논문에서는 영상의 상태가 좋지 않은 경우에도 자막 영역은 정확히 검출하였으나 개별문자 추출 단계에서 문자 인식을 위하여 영상을 개선하는 효과적인 방법에 대한 추가 연구가 필요하다.

향후 연구로는 압축 형태의 비디오 데이터를 처리할 수 있도록 하는 연구가 필요하며, 편집된 텍스트뿐만 아니라 씬 텍스트를 검출할 수 있도록 하는 것이 필요하다.

※ 참고문헌

[1] 구건서, "각인 문자 인식을 위한 특징 기반의 다중 연결 다층 인식기", 박사학위논문, 1997.

[2] 이종구, 양명섭, 이정열, 정찬근, 장옥배, "뉴스 비디오 검색을 위한 자동 인덱스 모델의 설계 및 구현," 한국정보과학회 인공지능연구회 97년도 춘계 학술 발표 논문집, pp. 103-112, 1997년 3월.

[3] 이미숙, 황본우, 이성환, "내용기반 비디오 검색을 위한 장면 전환 검출 방법의 성능 분석," 한국정보과학회 학술논문 발표집(B), pp. 543-546, 1997년 4월.

[4] 박영석, "문서화상의 표제영역에서 문자 추출", 정보과학회논문지(A) 제 24권 제 3호, pp, 221-231, 1997. 3월.

[5] Michael Philips, Wayne Wolf, "Video Segmentation Techniques for News," Proc. of SPIE Vol. 2916, pp. 243-251, 1996.

[6] D.Swanberg, C.F. Shu, and R. Jain, "Knowledge Guided Parsing in Video Databases," Proc. of SPIE'93 - Storage and Retrieval for Image and Video Database, Vol, 1908, pp. 13-24, 1993.

[7] Rakesh Mohan, "Text-based search of TV news stories", Proc of SPIE Vol, 2916, pp. 2-11, 1996.

[8] Hae-Kwang Kim, "Efficient Automatic Text Location Method and Content-Based Indexing and Structuring of Video Database", Academic Press, pp. 336 - 344, 1996.

[9 YU ZHONG, KALLEKARU and ANIL K. JAIN, "Locating Text in Complex Color Images", Pattern Recognition, Vol. 28, No. 10, pp. 1523-1535, 1995.

[10] H. Zhang, A. Kankanhalli and S. W. Smoliar, "Automatic Partitioning of Full-motion Video "Multimedia System, Vol. 1, No. 1, pp. 10 - 28, 1993.

[11] A. Hanjalic, R. L. Lagendijk, and J. Biemond. "A New Key-Frame Allocation Method for Representing Stored Video-Streams", Proc, of the First International Workshop on Image Databases and Multimedia Search, Armsterdam, The Netherlands, pp. 67 - 74, 1996.

[12] Y. Wu, and D. Suter, "A Comparison of Methods for Scene Change Detection in Noisy Image Sequence." Proc. of the Fist International conference on Visual Information Systems, Melbourne, Australia, pp. 459 - 468, 1996.

구 건 서

1997년 숭실대학교 대학원
공학박사
1996년~1997년 교육방송
(EBS) "컴퓨터는즐겁다"
진행
1999년~2000년
새정치국민회의,
21세기 지식정보화 추진
실무 위원
2002년~현재 중구
지역정보교육센터 소장
1993년~현재 숭의여자대학
인터넷 정보학과 교수
관심분야 : 영상처리,
문자인식, 전자상거래,
인터넷 응용