論文2003-40SP-1-2

뉴스 비디오 자막 추출 및 인식 기법에 관한 연구

(Study on News Video Character Extraction and Recognition)

金 鍾 烈 *, 金 聖 燮 **, 文 泳 植 *

(Jong Ryul Kim, Sung Sub Kim, and Young Shik Moon)

요 약

비디오 영상에 포함되어 있는 자막은 비디오의 내용을 함축적으로 표현하고 있기 때문에 비디오 색인 및 검색에 중요하게 사용될 수 있다. 본 논문에서는 뉴스 비디오로부터 폰트, 색상, 자막의 크기 등과 같은 사전 지식 없이도 자막을 효율적으로 추출하여 인식하는 방법을 제안한다. 문자 영역의 추출과정에서 문자영역은 뉴스 비디오의 여러 프레임에 걸쳐나 나오기 때문에 인접 프레임의 차영상을 통해서 동일한 자막 영역이 존재하는 프레임을 자동적으로 추출한 후, 이들의 시간적 평균영상을 만들어 인식에 사용함으로써 인식률을 향상한다. 또한, 평균 영상의 외각선 영상을 수평, 수직방향으로 투영한 값을 통해 문자 영역을 찾아 Region filling, K-means clustering을 적용하여 배경들을 완벽하게 제거함으로써 최종적인 자막 영상을 추출한다. 자막 인식과정에서는 문자 영역 추출과정에서 추출된 글자영상을 사용하여 white run, zero one transition과 같은 비교적 간단한 특징 값을 추출하여 이를 비교함으로써 인식과정을 수행한다. 제한된 방법을 다양한 뉴스 비디오에 적용하여 문자영역 추출 능력과 인식률을 측정한 결과 우수함을 확인하였다.

Abstract

Caption information in news videos can be very useful for video indexing and retrieval since it usually suggests or implies the contents of the video very well. In this paper, a new algorithm for extracting and recognizing characters from news video is proposed, without a priori knowledge such as font type, color, size of character. In the process of text region extraction, in order to improve the recognition rate for videos with complex background at low resolution, continuous frames with identical text regions are automatically detected to compose an average frame. The image of the averaged frame is projected to horizontal and vertical direction, and we apply region filling to remove backgrounds from the character. Then, K-means color clustering is applied to remove remaining backgrounds to produce the final text image. In the process of character recognition, simple features such as white run and zero-one transition from the center, are extracted from unknown characters. These feature are compared with the pre-composed character feature set to recognize the characters. Experimental results tested on various news videos show that the proposed method is superior in terms of caption extraction ability and character recognition rate.

Keyword : video indexing/retrieval, caption extraction, region filling K-means clustering, Character recognition

* 正會員, 漢陽大學校 컴퓨터工學科

(Department of Computer Science and Engineering, Hanyang University)

** 正會員, LG-OTIS SI 연구팀

(LG-OTIS R&D Center SI Team)

※ 본 연구는 한국과학재단 목적기초연구(R01-2000 000-00281-0)지원으로 수행되었음.

接受日字:2002年2月5日, 수정완료일:2002年11月29日

Ⅰ. 서 론

최근 멀티미디어 데이터 처리 기술의 급속한 성장과 고속 통신의 발전은 멀티미디어 데이터 서비스에 대한 높은 관심을 불러일으키고 있다. 여러 종류의 멀티미디어 데이터들 중에서도 비디오는 동영상과 함께 오디오와 자막 정보를 포함하고 있는 복잡한 성격의 데이터로서 그 중요성이 점차 증가하고 있으며 오락, 교육, 멀티미디어 애플리케이션 등의 넓은 분야에서 중요하게 사용되어지고 있다.

이렇게 방대한 크기와 양의 멀티미디어 데이터로부터 사용자가 원하는 데이터를 효율적으로 제공하기 위한 비디오 데이터의 자동 색인 기술이 요구되어 이를 위한 연구가 활발하게 진행 중이며 내용기반 동영상 검색 시스템은 이러한 기술의 대표적인 예이다. 내용기반 동영상 검색 기술은 추출된 영상의 특징 정보(컬러, 모양, 자막, 질감, 견본영상, 등)를 메타 데이터화하여 데이터베이스에 저장하고 이 특징에 대하여 질의함으로서 가장 유사한 영상을 추출하여 준다. 이들 특징 정보 중에서도 비디오 영상에 포함되어 있는 자막은 비디오의 내용을 함축적으로 표현하고 있기 때문에 이 자막을 정확하게 인식할 수 있다면 비디오 색인 및 검색에 중요하게 사용될 수 있다. 특히, 뉴스 비디오에 삽입되어 있는 자막 정보는 보도되고 있는 내용을 정확히 나타내며, 하이라이트가 되어있는 제목들은 보도 내용 전체를 대표하는 정보이다[1,2]. 따라서, 뉴스 자막 정보를 인식하여 색인 정보로 사용할 경우 사용자는 찾고자 하는 뉴스를 손쉽게 검색할 수 있다.

지금까지 자막을 포함하고 있는 프레임을 검출하기 위한 여러 가지의 시도들이 있어 왔다. 이들 방법들은 대부분 다음과 같은 문자의 성질을 사용하여 검출한다[3~5].

1. 문자들은 일정한 크기를 갖는다.
2. 문자들은 수평방향으로 일직선 나열되어 문자열을 이룬다.
3. 문자들은 배경과 대조된다.

기존에 제안된 텍스트프레임 검출 방법들로는 문자의 지형학적 문자 특성을 다단계(Multi level)로 추출함으로써 검출하는 방법[6]과 수평으로 나열되어 있는 문자열에서 나타나는 문자열의 주기적인 밝기의 차를 질감으로 검출하는 방법이 있다[7]. 문자영역의 질감을 사용한 방법으로는 문자영역이 배경에 대해 높은 밝기값의 차를 보인다는 특징(Gradient, Laplacian)을 이용한 방법[8] 등이 있고 또 최근에는 DCT 블록 내에서 수평, 수직, 대각선 경계를 찾는 방법들도 많이 사용되고 있다[9,10].

추출된 문자들을 인식하기 위해 기존의 상용 OCR (Optical Character Recognition)을 사용하여 인식하는 것은 비디오 텍스트 영상의 해상도가 낮고 복잡한 배경을 포함할 수 있기 때문에 매우 어려운 일이다. 비디오 텍스트 영상을 정확하게 인식하기 위해서는 복잡한 배경을 제거할 수 있는 영상 향상 과정이 필요하고, 비디오에서 동일한 텍스트 영상은 여러 프레임에 걸쳐 나타난다는 특징과 텍스트를 포함하는 동영상에서 배경은 대부분 움직이지만 텍스트의 이동은 거의 없다는 특징을 이용하여 몇 개의 연속된 프레임들을 논리합 연산을 수행함으로써 배경을 어느 정도 제거할 수 있는 방법들이 제안되었다[2].

배경과 잡음을 어느 정도 제거한 후에는 최종적으로 영상의 이진화 과정을 거쳐 문자영역과 배경영역을 분리하게 된다. 문자의 이진화 과정은 화소 밝기의 문턱치 값을 이용한 이진화 방법과, 영역 분할 및 합병(Region split and merge) 방법을 이용하여 추출하는 방법, 또 프레임간의 차를 이용한 문자 영역 추출 방법과 컬러의 축소(Color reduction)방법 등이 있다[6]. 하지만 이러한 방법들의 문제점은 문자 영역 추출을 위해서는 문자의 사전지식(문자의 크기, 색상, 서체 등)이 필요하다는 것이다.

본 논문에서는 동영상으로부터 자막을 가진 프레임을 자동으로 인식하고 텍스트 프레임으로부터 자막영역을 추출한 후, 이를 인식할 수 있는 효과적인 방법에 대하여 제안한다. 제안된 방법은 크게 세 가지 단계, 즉 텍스트 프레임 검출, 문자영역 추출, 문자인식 과정으로 나뉘어 진다. 먼저 1절 텍스트 프레임 검출과정에서는 동영상에서 텍스트를 포함하고 있는 프레임을 검출하기 위해 에지 영상을 구하여 이를 수평, 수직 방향으로 투영하여, 문자열 영역의 수, 위치, 크기, 분포 등을 계산하여 시간적으로 일정치 이상 유사하면 동일한 텍스트 프레임으로 간주하고 평균영상을 취한다. 2절 문자영역 추출과정에서는 시간적 평균 영상을 사용하여 Boundary region filling, Color clustering과 같은 다단

계 배경제거 및 검증 과정을 통해서 최종적으로 배경이 제거된 문자영상들을 획득한다. 3절 문자인식 과정에서는 인식될 문자로부터 비교적 단순한 White run, Zero-one transition from center 특징 값을 추출하여 폰트에 따라 이미 구성된 특징값과 서로 비교함으로써 문자인식 과정을 수행한다.

Ⅱ. 제안된 비디오 자막 인식 방법

1. 텍스트 프레임 검출

동영상에서의 문자는 다양한 색상, 서체, 크기 등을 갖기 때문에 문자영역의 유무를 일반화 하기는 쉬운 일이 아니다. 하지만 동영상에서의 문자는 여러 프레임에 걸쳐 나오기 때문에 이러한 특성은 문자 프레임 검출에 유용하게 사용될 수 있다.

제안된 동영상 텍스트 프레임 검출방법에서는 먼저 동영상 텍스트 샷(shot)을 찾는다. 동영상 텍스트 샷이란 동영상 내에서 동일한 텍스트를 갖는 연속된 프레임을 말한다. 이러한 텍스트 샷은 문자영역의 분포가 갑작스럽게 변하는 프레임을 찾아 텍스트 샷의 시작과 끝 프레임으로 검출하고 동영상 텍스트 샷에 존재하는 모든 프레임을 문자영역 추출 시 최대한 이용한다. 다음의 <그림 1>은 텍스트 샷을 검출하기 위한 전체적인 과정을 나타낸다.

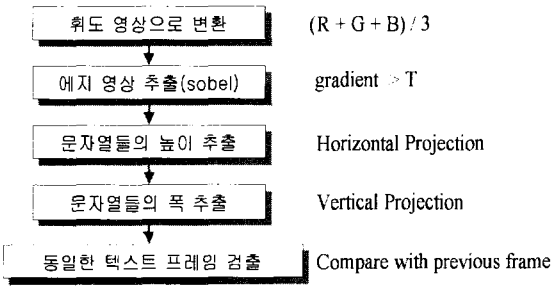

그림 1. 동영상의 텍스트 샷 검출 과정
Fig. 1. Text shot extraction in videos.

텍스트 프레임을 찾기 위해서는 먼저 각 프레임에서 후보 문자영역을 추출해야 한다. 이를 위해서 동영상의 각 프레임을 컬러 영상을 휘도(gray-scale) 영상으로 변환한 후, Sobel 연산자를 사용하여 에지 영상을 만든다. 에지 영상으로부터 수평방향으로 투영, 수직 방향으로 투영시켜 분포가 조밀한 부분의 시작과 끝을 찾아 각 문자열의 높이를 구한다. 텍스트 영역에서 문자들은 수평으로 나열되어 있기 때문에 영상에서 수평방향으로 넓은 후보 텍스트 영역을 갖는 프레임을 텍스트 프레임으로 검출한다.

후보 문자 영역이 결정되면 현재 프레임과 이전 프레임간의 문자 영역의 수, 위치, 크기, 분포 등을 비교하여 일정치 이상 유사하면 동일한 텍스트 프레임으로 간주되고 이는 하나의 텍스트 샷이 된다. 다음의 <그림 2>는 텍스트 샷을 찾아서 문자 영역을 찾기 위해 샷에 포함되어 있는 모든 프레임을 평균한 영상이다.

그림 2. 동영상 텍스트 샷 내에 포함된 프레임들의 평균 영상
Fig. 2. The averaged frame in video text shot

Ⅱ. 문자영역 추출

제안된 문자영역 추출 방법에서는 문자영역 추출을 위해 먼저 동일한 텍스트가 나타나는 프레임들의 시간적 평균을 통해 영상의 화질을 향상하고, 영상에 내에 존재하는 문자의 영역들은 3단계의 배경제거 과정을 거쳐 문자영역을 추출하게 된다. 1차 배경 제거 과정은 문자영역의 외각선 상에 놓여있는 화소들의 컬러 값을 초기값(seed)으로 사용하여 Region filling을 수행한다.

배경이 어느 정도 제거된 글자 영상으로부터 각 글자 영역의 분산 값을 구하고, 이를 토대로 1차 배경제거의 결과를 검증하여 추가적인 2차 배경 제거 과정을 적용할지를 결정한다. 2차 배경 제거 과정은 1차 배경 제거 결과에 따라 K-means color clustering을 추가적으로 수행한다. 마지막으로 3차 배경제거는 크기가 작은 잡음 등을 제거하여 문자영역 추출을 완료한다. 다음의 <그림 3>은 본 논문에서 제안된 문자 영역 추출 방법의 전체적인 과정을 보여주고 있다.

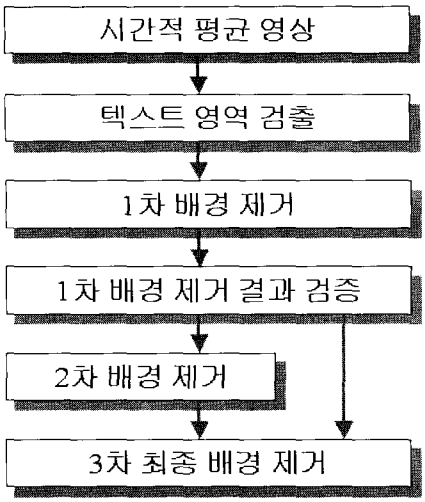

그림 3. 문자영역 추출 과정
Fig. 3. Text region extraction.

1. 문자 영상 향상

동영상의 화질이 저하되어 있거나 복잡한 배경이 있는 동영상에서는 하나의 프레임으로부터 텍스트를 추출하는 것보다는 동일한 텍스트를 갖는 모든 프레임을 사용한다면 보다 좋은 텍스트 영역 추출 결과를 얻을 수 있다. 비디오 텍스트의 변환이 일어나는 부분의 시작과 끝을 찾으면 유사성이 있는 연속된 텍스트 프레임들의 집합을 텍스트 샷으로 결정하고, 텍스트 샷 사이에 있는 모든 프레임들의 정보를 텍스트영역 추출 과정에 사용 할 수 있게 된다.

이를 위해 앞에서 설명한 텍스트 프레임 검출 방법을 사용하여 동일한 텍스트의 처음과 마지막 프레임을 찾는다. 비디오에서 배경은 대부분 움직이지만 동일한 텍스트는 여러 프레임에 걸쳐 변화가 없다는 특징을 이용하여 텍스트 샷에 존재하는 모든 프레임의 시간적 평균 프레임을 구한다. 시간적 평균 프레임에서 배경부분은 대부분 변화하기 때문에 배경의 움직임이 많을수록 컬러에 변화가 많이 일어나는 반면에 텍스트영역의 컬러는 적은 변화만 일어나게 된다. <그림 4>는 MPEG 비디오에서 하나의 텍스트 샷에 존재하는 모든 I 프레임들의 시간적 평균프레임을 만들어 영상의 질을 향상시키고 프레임의 평균 영상에 나타나는 문자영역을 찾은 결과이다.

2. 배경 제거

찾아진 문자영역의 외각선 상에 놓여있는 화소들의 컬러 값을 초기값(seed)으로 하여 Region filling을 수행함으로써 경계와 유사한 색상을 갖는 부분들을 제거한다. 식 (1)은 Region Filling을 위한 두 컬러의 거리를 결정하는 식이다.

$$dist = (R_1 - R_2)^2 + (G_1 - G_2)^2 + (B_1 - B_2)^2 \quad (1)$$

이때 R_1, G_1, B_1은 초기의 컬러 값을 나타내고, R_2, G_2, B_2는 문자영역 내 임의의 화소의 컬러 값이다. 다음의 <그림 5>는 1차 배경제거를 수행하여 얻어진 결

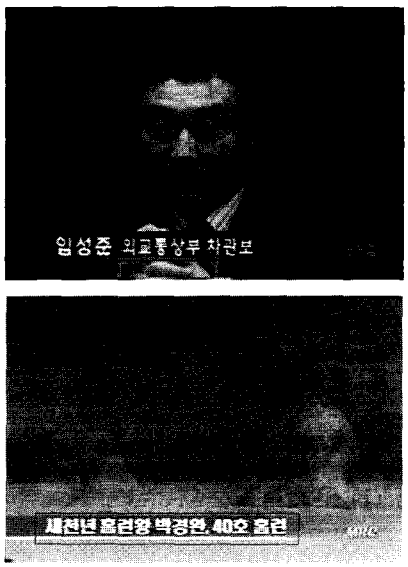

그림 4. 찾아진 문자영역
Fig. 4. The detected text region.

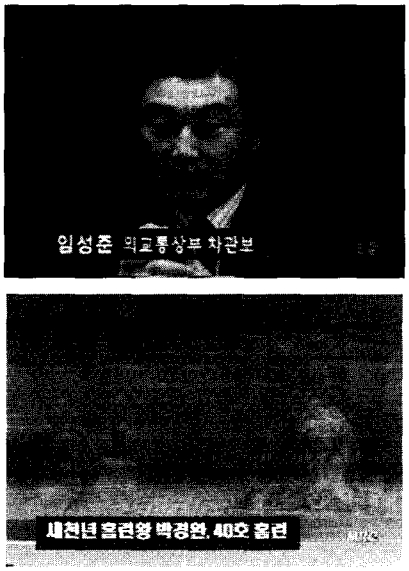

그림 5. Region filling을 사용하여 배경을 제거한 결과
Fig. 5. background removal using region filling.

과 영상이다.

(3) 1차 배경 제거의 결과 검증

1차 배경 제거 단계를 거쳐 어느 정도 분리된 각각의 글자영역의 분산값를 구하여 1차 배경제거의 결과를 검증한다. 1차 배경 제거를 수행 후, 이를 검증하는 이유는 Region filling 만으로도 대략적인 배경 제거를 할 수 있지만 글자 주위의 제거되지 않은 배경들이 남아 있거나 'ㅁ,ㅇ,ㅂ' 등과 같은 글자에서 나타나는 고립된 영역은 제거되지 않는 결과가 발생하기 때문이다. 이를 위해 먼저 한 글자 영역의 전체 분산값을 구한다. 만약 1차 단계에서 글자의 분리가 잘 되었다면 동일한 글자영역에서의 분산값은 작은 값을 갖게 되고 이런 경우에는 1차 배경 제거 과정만을 수행하고, 분리가 잘 되지 않았을 경우에는 큰 분산값을 갖게 되며 이런 경우에는 추가적인 2차 배경 제거 과정을 수행한다. 제안된 방법에서는 k-means color clustering을 통해 글자 영역을 두개의 cluster로 나눔으로써 글자와 배경을 최종 분리할 수 있는 2차 배경제거 과정을 수행한다.

하지만 1차 배경 제거 단계에서 이미 배경제거가 잘 되어진 글자들에 대해 color clustering을 적용하면 오히려 글씨의 획이 사라지는 등의 좋지 않은 결과를 초래하기 때문에 color clustering에 앞서 1차 배경제거 결과에 따라 2차 배경 제거 과정 적용의 필요성을 검증하여야 한다. 1차 배경 제거의 결과 검증을 위한 상세 과정은 다음과 같다. 여기서 T_1, T_2와 T_e는 실험적으로 결정되는 문턱치이다.

1. Otsu의 이진화 방법을 사용하여 경계 'K'를 구한다.
2. 글자영역 전체의 표준편차 V_e를 구한다.
3. 'K'를 기준으로 나누어지는 2개의 영역의 표준편차 V_1, V_2를 구한다.
4. if $((V1 > V_2 * T_1 \| V_2 > V_1 * T_2)$ && $V_e < T_e)$
   2차 배경제거 과정 생략.
   else
   2차 배경제거 과정 수행

<그림 6(a)>는 region filling을 수행한 후 모든 글자 영역에 대해 color clustering을 했을 때의 결과이다. <그림 6(a)>에서 보는바와 같이 동그라미가 쳐져 있는 부분의 획이 심하게 상해 있는 것을 볼 수 있다. 하지만 <그림 6(b)>에서는 각 글자 영역별로 1차 배경제거 과정의 결과를 검증한 후, 추가적이 과정을 필요로 하는 글자 영역에만 color clustering을 수행한 결과로 이는 획이 손상되는 문제를 해결할 수 있음을 확인할 수 있다.

그림 6. (a) 모든 문자에 2차 배경 제거과정을 수행한 결과

Fig. 6. (a) Result of applying 2nd background elimination step to all characters.

그림 6. (b) 필요한 문자에 대해서 2차 배경 제거과정을 수행한 결과

Fig. 6. (b) Result of applying 2nd background elimination step to the selected characters.

4. Color clustering을 이용한 2차 배경의 제거

1차 배경 제거 과정에서 추출된 문자 영역이 높은 분산값을 갖는 경우, 즉 2차 배경 제거 과정이 필요한 경우에는 color clustering을 통해 글자 영역과 남아있는 배경영역을 분리한다. Clustering의 입력벡터는 각 화소의 컬러 값이고 글자와 배경의 2개의 cluster를 갖는 k-means algorithm을 사용한다. 보다 나은 clustering을 위해 글자영역의 컬러 히스토그램(8x8x8)에서 나타나는 두개의 local max color를 찾아 각 cluster의 중심값으로 한다.

Clustering을 마친 후, 두개의 cluster중 많은 수의 요소(element)를 갖으며 상대적으로 밝은 밝기값을 갖는 cluster를 선택하여 글자 영역으로 하고 적은 수의 요소(element)와 상대적으로 어두운 부분을 배경으로 선택한다. <그림 7>은 2차 배경 제거 단계를 수행한 결과이다.

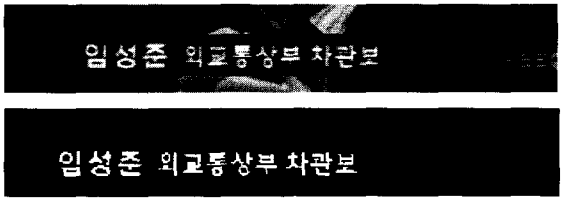

그림 7. 2차 배경 제거 단계를 수행한 결과

Fig. 7. Result of applying 2nd background elimination step.

Ⅲ. 문자 인식

비디오는 다양한 서체의 문자들을 포함하고 있고 비디오 문자영역 분할 시 서체의 형태의 변화가 일어나기 때문에 일반적인 문자인식 기법으로 이를 인식하기는 매우 힘들다. 본 논문에서는 다양한 서체의 문자를 인식하기 위해 비디오에서 직접 여러 서체의 문자들로부터 특징값을 추출하여 사용한다. 문자의 인식은 사전 추출된 각 문자들의 특징값과 인식하고자 하는 문자의 특징값과 비교를 통해 인식과정을 수행한다.

1. 특징값 추출

특징값 추출을 위해서는 <그림 8>에서 보는바와 같이 투영을 이용하여 배경이 모두 제거된 문자영상으로부터 각 글자의 영역을 얻어온다. 비디오에는 다양한 크기의 문자가 나타나기 때문에 인식 과정에 앞서 일정한 크기(30×30)로 정규화 시킨다.

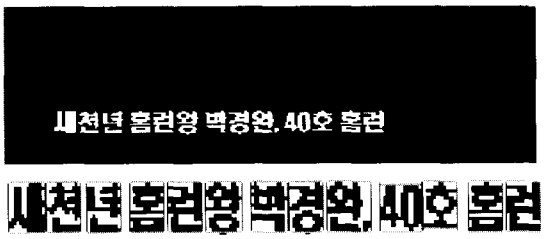

그림 8. 각각의 글자 영역 분할 및 정규화
Fig. 8. Region segmentation and normalization for each character.

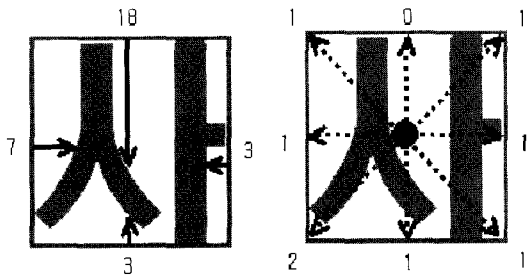

그림 9. (a) White run. (b) Zero-one transition from center
Fig. 9. (a) White run. (b) Zero-one transition from center.

인식을 위한 특징값은 각 글자들의 특성을 독자적으로 표현하여야 하며 많은 수의 문자와 서체를 표현해야 하기 때문에 가능한 적은 수의 특징값을 갖는 것이 바람직하다. 제안된 문자인식 방법에서는 <그림 9>에서 각각 보는바와 같이 글자의 상, 하, 좌, 우로부터 외각 정보를 나타내는 white run과 글자영역 중심으로부터 외각 쪽으로 획의 분포를 나타내는 zero-one transition을 특징값으로 사용한다. 사용된 특징 값들은 비교적 단순한 형태로 보이나 엄청난 양의 글자 수와 각 글자마다 가질 수 있는 폰트 등을 고려해서 비교적 간단하면서도 인식 성능을 높일 수 있는 특징 값이다.

2. 특징값의 비교

앞에서 설명한 특징값들을 사용하여 모든 문자에 대한 특징값의 집합을 만든다. 문자의 인식은 모든 문자의 특징값의 집합과 인식하고자 하는 문자와의 비교를 통해 가장 유사한 특징값을 갖는 글자를 찾아냄으로써 인식을 수행한다. 비디오에서는 다양한 서체의 문자들이 나타나기 때문에 어느 특정 서체의 문자들의 특징값만을 사용하여 인식을 수행하기는 힘들다. 따라서 가능한 많은 수의 서체를 사용하여 특징값의 집합을 구성하면 높은 인식률을 기대할 수 있다. 하지만 너무 많은 수의 서체를 사용하게 되면 특징값 비교를 위해 많은 시간이 소요됨으로 적절한 서체와 글자의 수를 정하는 것이 중요하다. 식 (2)는 인식하고자 하는 글자와 특징값의 집합들과의 비교를 위한 식이다.

$$best\ \ match = \min_{\alpha}(\min_{\beta}(\sum_{i=0}^{N}\omega_i(d_{\alpha\beta\ i}-u_i)^2)) \quad (2)$$

여기서 α는 서체의 집합이고 β는 각 서체별 글자의 집합이다. N은 각 문자를 인식하기 위해 사용된 특징값의 수이고 ω_i는 각 특징값에 대한 가중치이다. $d_{\alpha\beta i}$와 u_i는 각각 특징값 집합에 있는 α서체 β글의 i번째 특징값과 인식하고자 하는 문자의 i번째 특징 값이다. 다음의 <그림 10>은 white run과 zero-one transition from center 특징 값을 사용하여 인식을 수

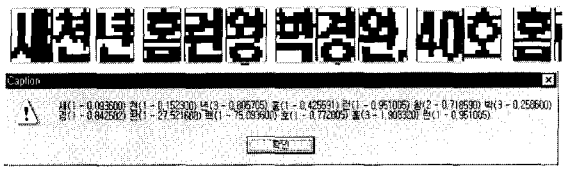

그림 10. 문자인식 수행 결과
Fig. 10. Character recognition result

행한 결과이다. 결과 화면에서 괄호 안에 있는 값은 결정된 폰트종류와 그 때의 차이 값을 나타낸다.

Ⅲ. 실험 결과 및 성능 분석

본 논문에서는 KBS, MBC, SBS 뉴스 및 스포츠 뉴스를 MPEG-1 동영상으로 코딩하여 제안된 문자영역 추출과 인식 기법을 평가하였다. 문자인식을 위해서는 동영상으로부터 직접 문자를 추출하여 가장 많이 사용되는 견고딕, 굴림, 둥근고딕, 휴먼고딕 견명조의 5개의 폰트에 대해서 폰트당 588개의 문자집합을 생성하였다. 이는 588자 만으로도 일반적으로 비디오에 나타나는 자막의 98% 이상을 포함하기 때문이다. 성능평가를 위해서는 텍스트 프레임 추출 능력과 문자 인식률에 대하여 결과를 도출한다.

1. 실험 환경

본 논문에서 동영상 자막영역 추출 및 인식을 위해서 제안된 시스템은 IBM Pentium Ⅲ 800 MHz 컴퓨터를 사용하였고 프로그램 언어는 Visual C++ 6.0을 사용하였다. 실험에 사용된 뉴스 동영상은 다음의 <표 1>과 같이 KBS, MBC, SBS 뉴스 및 스포츠 뉴스를 MPEG-1 동영상으로 코딩한 6시간 분량이다.

표 1. 실험에 사용된 뉴스 동영상들
Table 1. News videos used in the experiment

동영상의 종류	시간	텍스트 프레임의 갯수	문자의 갯수
KBS	2시간 36분	397	4822
MBC	1시간 54분	243	3170
SBS	1시간 30분	236	3082

그림 11. 뉴스 비디오 샘플들
Fig. 11. News video samples.

<그림 11>은 실험에 사용된 뉴스 동영상들의 샘플들을 보여주고 있다.

2. 실험결과 및 성능 평가

다음의 <그림 12>는 본 논문에서 제안된 방법으로 자막추출부터 인식을 수행한 전체적인 과정에 대한 결과이다.

실험 결과에 대한 성능 평가를 위해서는 다음과 같은 성능 평가 기준을 설정하였다. 첫 번째는 텍스트 프레임의 검출 능력 TFER(Text Frame Extraction Rate)과 오 검출 비율 EER(Erroneous Extraction Rate)이다. 오 검출은 텍스트 프레임인데 찾지 못하는 경우 FTF(Fail Text Frame)와 텍스트 프레임이 아닌 것을 텍스트 프레임으로 검출하는 경우 NTF(No Text Frame)를 합한 값이다.

$$TFER = ETTN / TTN$$
$$EER = (FTF + NTF) / TTN \quad (3)$$

여기서 ETTN(Extracted Total Text Number)은 추출된 총 텍스트 프레임의 개수이고, TTN(Total Text Number)은 실험에 사용된 동영상 내에 존재하는 텍스트 프레임의 개수이다.

위에서 제시한 성능 평가 기준에 대한 종합적인 결과는 <표 2>에서와 같다.

표 2. 텍스트 프레임 검출 결과
Table 2. The result of text frame extraction.

전체 텍스트 프레임 개수	TFER	EER		
		FTF	NTF	Total
876	830개	27개	19개	46개
	94.8%	3.1%	2.1%	5.2%

두 번째 성능평가 기준은 문자 인식률 CRR (Character Recognition Rate)이다. 인식률에서는 두 가지 경우를 대상으로 실험 결과를 도출한다. 먼저 텍스

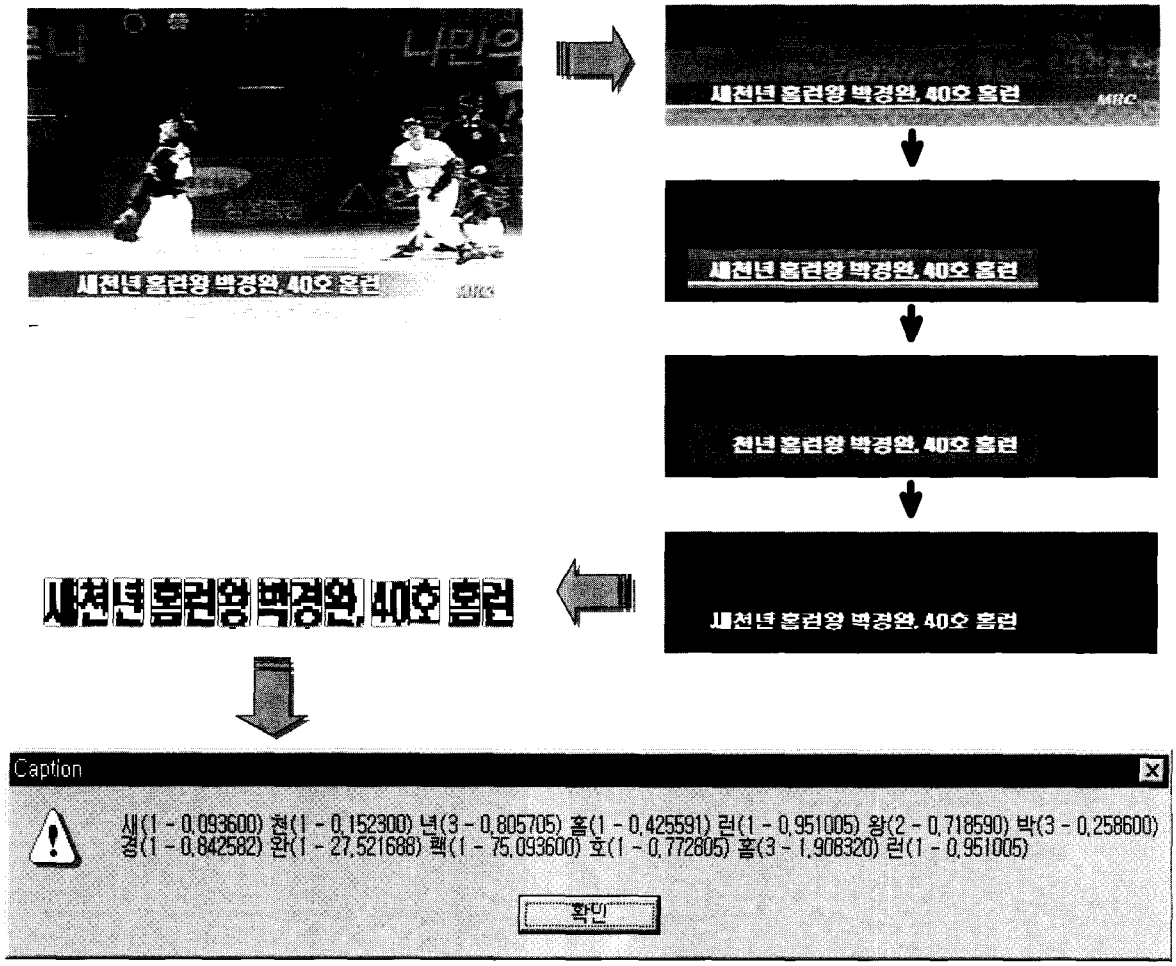

그림 12. 실험 결과
Fig. 12. Experimental results.

트 프레임 검출 능력과 인식능력을 동시에 고려한 결과와, 문자인식 능력만을 고려한 문자 인식률에 대한 결과를 보면 <표 3>과 같다.

표 3. 문자인식률
Table 3. Character recognition rate.

	텍스트 프레임 개수	문자의 총 개수	인식된 문자 개수	인식률
TTN인 경우	876개	11,074	8,460	76.4%
ETTN인 경우	830개	10,298	8,460	82%

결과에서 보는 바와 같이 시스템의 전체의 문자 인식률은 76.4%의 인식률을 보였고, 텍스트 프레임이 정확하게 검출되었다는 가정 아래에서의 인식률 즉, 제안된 문자 인식기의 인식률은 82%의 인식률을 보였다.

실험에 사용한 뉴스 동영상을 MPEG-1으로 코딩시 화질 저하 문제가 생기고, 글자의 크기가 너무 작아서 인식이 불가능한 문제도 발생하여 전체적인 인식률을 하락시키는 결과를 볼 수 있었다.

Ⅳ. 결론 및 향후 연구 과제

본 논문에서는 글자의 색상, 크기, 서체 등의 사전 지식 없이도 비디오로부터 문자 영역을 추출하는 방법을 제안하였다. 기존의 동일한 자막 프레임을 판별하는 방법을 보완하여 시작 프레임과 끝 프레임을 찾아서 이를 텍스트 샷이라 하고, 이들 사이에 존재하는 모든 프레임을 이용하여 문자 영역 분할에 사용하였다. 3단계

에 걸친 배경제거 및 검증을 통해 뛰어난 문자 영역 이진화를 수행 할 수 있었다. 또한 문자 인식 과정에서는 동영상으로부터 다양한 폰트의 문자들을 직접 추출하여 인식 과정에서 사용하였고 비교적 단순한 특징값을 사용하여 인식에 필요한 시간을 최소화하였다. 실험 결과에서는 텍스트 프레임 검출 능력은 95% 정도의 검출 능력을 보였고, 검출된 텍스트 프레임에서 대한 인식률은 82% 정도의 문자를 인식하는 것으로 나타났다. 이러한 결과는 MPEG-2와 같은 고화질 동영상을 사용할 경우, 보다 높은 인식률을 보일 것으로 생각된다. 향후 연구 과제로는 1차 배경 제거 단계의 검증시 사용되는 분산값을 자동으로 구하는 문제와 문자 인식기의 성능을 향상시켜 인식률을 올리는 것이다.

참 고 문 헌

[1] 최경주, 변혜란, 이일병, "이진화를 위한 영상 강화 기법에 관한 연구," 제 10회 영상처리 및 이해에 관한 워크샵 발표 논문집, pp. 176~181, 1998

[2] 곽상신, 김소명, 최영우, 정규식, "효율적인 비디오 자막 인식을 위한 영상 향상 방법," 제 12회 영상처리 및 이해에 관한 워크샵 발표 논문집, pp. 342~347, 2000

[3] U. Gargi, S. Antani, R. Kasturi, "Indexing Text Event in Digital Video Database," Proc. 14th International Conference of Pattern Recognition, pp. 916~919, 1998.

[4] A. K. Jain and B. Yu, "Automatic Text Location in Images and Video Frames," Pattern Recognition, Vol. 31, No. 12, pp. 2,055~2,075, 1998.

[5] S. W. Lee, D. J. Lee, H. S. Park, "A New Methodology for Grayscale Character Segmentation and Recognition," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 18, No. 10, pp. 1,045~1,050, 1996.

[6] 전병태, 배영래, 김태윤, "일반화된 문자 및 비디오 자막 영역 추출 방법," 정보과학회 논문지 : 소프트웨어 및 응용 제 27권 제 6호, pp. 632~641, 2000

[7] A. K. Jain, Y Zhaong, "Page Segmentation Using Texture Analysis," Pattern Recognition, Vol. 29, No. 5, pp. 743~770, 1996.

[8] H. Kuwano, Y. Taniguchi, H. Arai, "Telop-on-Demand: Video Structuring and Retrieval Based on Text Recognition," IEEE International Conf. Multimedia and Expo, 759~762, 2000.

[9] Y. Zhong, H. Zhang, A. K. Jin, "Automatic Caption Localization in Compressed Video," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 22, No. 4, pp. 385~392, 2000.

[10] 박영규, 김성국, 유원영, 김준철, 이준환, "MPEG II 뉴스 영상에서의 문자영역 추출 및 문자인식," 신호처리합동 학술대회 논문집, pp. 117~120, 1998

저 자 소 개

金 鍾 烈(學生會員)

1997년 2월 한양대학교 전자계산학과 졸업(학사). 1999년 2월 한양대학교 대학원 전자계산학과 졸업(석사). 1999년 3월~현재 한양대학교 대학원 전자계산학과 박사과정. <주관심분야 : 영상처리, 컴퓨터 비전, 패턴인식, 멀티미디어 정보보호 등>

金 聖 燮(正會員)

1998년 8월 Univ. of Nevada 전자계산학과 졸업(학사). 2001년 2월 한양대학교 대학원 전자계산학과 졸업(석사). 2001년 3월~현재 LG-OTIS SI 연구팀 연구원. <주관심분야 : 영상처리, 패턴인식 등>

文 泳 植(正會員)

1980년 2월 서울대학교 공과대학 전자공학과 졸업(학사). 1982년 2월 한국과학기술원 전기 및 전자공학과 졸업(석사). 1990년 University of California at Irvine Dept. of Electrical and Computer Engr. (박사). 1982년~1985년 한국전자통신연구소 연구원. 1989년~1990 Inno Vision Medical 선임연구원. 1990년~1992년 생산기술연구원 선임연구원. 1992년~현재 한양대학교 전자계산학과 부교수. <주관심분야 : 영상처리, 컴퓨터 비전, 패턴인식 등>