뉴스 비디오의 내용기반 검색을 위한 자동 인덱싱

양 명 섭† · 유 철 중†† · 장 옥 배†††

요 약

본 논문은 내용에 기반한 뉴스 비디오의 인덱싱과 검색을 위한 통합된 해결책을 제안한다. 현재 일반적인 비디오의 자동 인덱싱은 불가능하지만 뉴스 비디오와 같은 구조가 명확한 경우는 가능하다. 이러한 뉴스의 구조화된 지식을 이용하여 키 프레임들을 자동 추출하기 위해서 제안된 우리의 모델은 뉴스사건 분할, 자막 인식, 검색 브라우저 모듈로 구성되어 있다. 첫 번째로 뉴스사건의 분할 모듈은 얼굴인식에 기반하여 사건의 중심인 앵커 장면을 인식하고 앵커 장면의 공간적 정보를 이용하여 뉴스사건을 분할한다. 다음으로 뉴스아이콘을 추출한다. 자막인식 모듈은 먼저 자막의 특성을 이용하여 자막 프레임을 검출하고 분리병합 방법을 이용하여 문자열을 추출한다. 다음으로 문자인식기(OCR)를 이용하여 문자인식을 한다. 마지막으로 검색 브라우저 모듈은 다양한 검색 방법이 가능하도록 하였다.

Automatic Indexing for the Content-based Retrieval of News Video

Myung-Sup Yang† · Cheol-Jung Yoo†† · Ok-Bae Chang†††

ABSTRACT

This paper presents an integrated solution for the content-based news video indexing and the retrieval. Currently, it is impossible to automatically index a general video, but we can index a specific structural video such as news videos.

Our proposed model extracts automatically the key frames by using the structured knowledge of news and consists of the news item segmentation, caption recognition and search browser modules. We present above three modules in the following: the news event segmentation module recognizes an anchor-person shot based on face recognition, and then its news event are divided by the anchor-person's frame information. The caption recognition module detects the caption-frames with the caption characteristics, extracts their character region by the using split-merge method, and then recognizes characters with OCR software. Finally, the search browser module could make a various of searching mechanism possible.

1. 서 론

디지털 비디오가 현대의 중요한 정보매체라는 것은 의심할 여지도 없으며 정보통신의 발달과 더불어 멀티미디어 정보 서비스에 대한 높은 관심을 불러일으키고 있다. 비디오 정보는 선형 또는 공간 자료 구조를 갖고 있으며 비정형화된 자료의 형태로 존재한다. 또한 자료의 종류가 다양하고 방대하며 기존의 데이터 모델링 기법으로는 효율적으로 표현할 수 없는 복잡하고 다양한 관계성을 포함하고 있다. 따라서, 비디오 정보를 검색하기 위해서는 기존의 텍스트 기반의 정보 검색 시스템과는 다른 접근 방법이 필요하다.

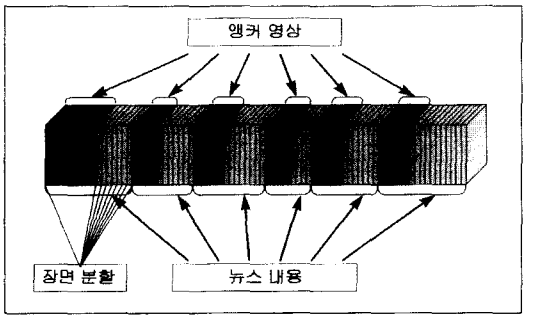
† 준 회 원 : 전북대학교 컴퓨터과학과 박사과정
†† 종신회원 : 전북대학교 컴퓨터과학과 교수
††† 정 회 원 : 전북대학교 컴퓨터과학과

논문접수 : 1997년 11월 27일, 심사완료 : 1998년 2월 25일

현재 멀티미디어 정보 검색에 관한 연구들은 극히 초보적인 단계로서 단순 속성이나 수동식 기술, 또는 캡션 자료를 이용한 텍스트의 매칭에 의한 경우가 대부분이다. 그러나 멀티미디어 정보는 여러 가지 속성을 지니고 있으며, 모든 멀티미디어 정보에 대한 기술을 사람이 수행해야 할 때는 엄청난 작업량이 수반될 뿐 아니라 동일한 데이터에 대한 기술이 주관에 따라 달라질 수도 있게 된다. 이러한 문제점들 때문에 멀티미디어 정보의 검색을 위해서는 데이터의 내용에 의해 관련 정보를 검색하는 내용기반(Content-based)의 검색 기술에 대한 연구가 필수적이다[4, 7, 9].

비디오 검색을 위한 국내외 연구 동향을 살펴보면 주로 비디오 분할에 관한 연구가 많이 이루어져 있다. 또한 범용 비디오의 효율적인 검색을 위한 많은 연구가 진행되고 있지만 아직은 영상과 음성인식 기술의 한계로 자동 인덱싱은 불가능하다. 그러나 뉴스비디오와 같은 비디오 구조가 명확한 경우에는 가능하다. 국내에서는 일부 대학과 연구소에서 연구가 이루어지고 있으나 아직은 미비한 상태이다.

뉴스 비디오는 (그림 1)과 같이 시간적으로는 앵커 장면을 중심으로 뉴스사건이 반복적으로 구성되어 있으며, 공간적으로는 각 장면에 뉴스사건을 대표하는 뉴스 아이콘 자막문자 정보를 가지고 있다. 본 논문에서는 이러한 뉴스 비디오의 사전 지식을 이용하여 뉴스의 대표영상인 키 프레임을 자동 추출하여 인덱싱하고 검색하는 방법을 제시한다. 제안된 인덱싱 모델은 다른 연구 방법과 달리 얼굴인식에 기반하여 뉴스사건을 분할하고 뉴스 아이콘을 추출 생성함으로써 아이콘 기반 검색이 가능하도록 하였다. 또한 자막문자 검출과 인식과정을 도입하여 내용기반 검색도 가능하도록 하였다.



(그림 37) 뉴스비디오 구조
(Fig. 1) News video structure

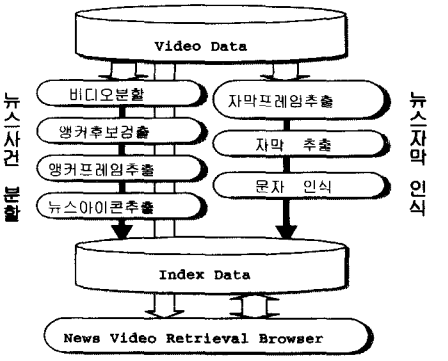
본 시스템의 구현은 Windows-NT가 탑재된 컴퓨터를 기반으로 인터넷과 연동하여 주문형 뉴스에 응용이 가능하도록 설계 및 구현되었다.

본 논문의 구성은 다음과 같다. 2장에서는 뉴스 비디오 인덱싱 방법에 대해 기술하며, 3장에서는 검색 브라우저에 대해 기술한다. 4장에서는 제안된 방법의 실험 결과를 보여주고, 마지막으로 5장에서는 결론을 맺는다.

2. 뉴스 비디오 인덱싱

Zhang[7] 등이 제안한 방법은 단순히 앵커모델에 따른 뉴스 사건의 분할에 국한하여 검색하는 방법을 제안하였다. 이와는 달리 제안된 우리의 모델은 앵커인식에 기반하여 뉴스사건을 분할한다. 또한 뉴스사건을 대표하는 뉴스 아이콘의 추출과 부가적인 정보를 가지는 자막문자 인식과정을 거쳐 아이콘과 자막 기반 검색이 가능하도록 설계한다. 방송 뉴스 프로그램은 앵커와 뉴스 아이콘을 포함하는 앵커 장면(anchor shot)의 공간적인 구조, 장면과 뉴스사건들 사이의 시간적인 구조를 가지는 정형화된 모델을 따른다(그림 1).

본 논문은 이러한 구조적 지식을 이용하며 뉴스 비디오 인덱싱 과정은 (그림 2)과 같이 뉴스 사건 분할에 의한 아이콘 추출과 자막 문자 정보를 추출하여 인식하는 과정으로 이루어진다[9].



(그림 38) 모델의 인덱싱 과정
(Fig. 2) Process to index model

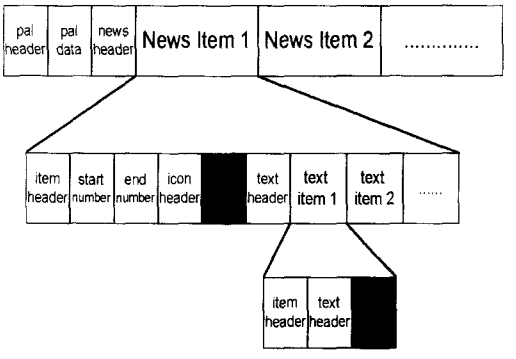
뉴스 아이콘과 자막문자 영역을 추출하기 위한 과정은 첫째, 프레임들 사이의 장면 변화를 검출한다. 프레임들 사이의 장면 변화를 검출하기 위한 방법으로 개선된 화소 비교법과 히스토그램 비교법이 사용되었다. 프

레임들의 장면 변화 검출 부분에서 시작 프레임의 번호와 끝 프레임의 번호를 얻어낼 수 있다. 둘째, 뉴스 사건의 분할과 공간적 정보를 가지는 뉴스 아이콘을 추출한다. 아이콘을 추출하기 위해서는 장면 변화로 검출된 프레임에서 색상 정보와 모양 정보를 이용하여 앵커 영역을 검출하며 검증과정을 거친 후 뉴스 아이콘을 추출한다. 아이콘 추출 부분에서는 아이콘의 자막 수와 뉴스 아이템의 아이콘을 얻을 수 있다. 셋째, 자막문자 영역의 추출과 인식을 한다. 뉴스의 자막문자 영역 추출 방법은 장면 변화로 검출된 프레임에서 색상 정보를 이용하여 추출한다. 자막은 자막의 배경선 색상 정보와 문자의 색상정보와 경계선 정보를 이용한다. 자막 문자 영역 추출 부분에서는 뉴스 자막 수를 얻을 수 있다. 추출된 뉴스 아이콘과 자막문자는 뉴스 검색 브라우저에게 중요한 요소로 제공된다. 뉴스 비디오의 인덱싱 요소는 다음과 같다(그림 3).

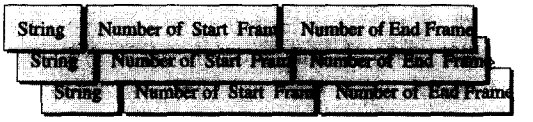
- 뉴스 사건의 개수
- 뉴스 사건의 시작과 종료 프레임 번호
- 뉴스 사건에 포함된 뉴스 아이콘과 자막 이미지
- 뉴스 사건에 포함된 자막의 개수

또한, 인식된 문자열로 인덱싱하는 구성 요소로는 (그림 4)와 같이 인식된 문자열과, 시작 프레임 번호, 종료 프레임 번호로 이루어진다.

이러한 정보들 중에서 뉴스 아이콘과 뉴스 자막은 프레임의 부분적인 이미지이며 뉴스를 이해하는 중요한 정보를 제공한다. 추출된 이미지는 인덱스 파일에 이미지 형태와 문자열로 저장하며 그 외의 정보들은 수치값으로 저장한다.



(그림 39) 뉴스사건 인덱스 파일 구조
(Fig. 3) Structure of news event index file



(그림 40) 뉴스자막 인덱스 파일 구조
(Fig. 4) Structure of news caption index file

2.1 비디오 분할

디지털 비디오에서 비디오의 분할[5,8,10]과 내용을 포함하는 물체의 추출은 인덱싱의 가장 기본이 되는 방법 중의 하나이다. 비디오 분할은 각각 하나의 내용을 표현하는 가장 작은 단위를 인식하기 위해서 사용된다. 일반적으로 비디오를 분할하기 위해서는 제일 먼저 비디오를 컷(cut) 단위로 나누며 컷을 검출하는 기법으로는 화소의 색상 또는 세기를 이용하는 히스토그램 비교(histogram difference)방법들을 이용한다[1, 2]. 이 방법은 화소의 세기(Y) 성분과 색상(Cb, Cr) 성분을 히스토그램으로 표현하여 식 (1)과 같이 히스토그램 차로 유사도를 측정하며 화소단위의 비교보다 카메라 또는 물체의 움직임에 덜 민감하다.

$$SD_i = \sum_{j=1}^{G} | H_i(j) - H_{i+1}(j) | \qquad (1)$$

j : G개의 밝기 레벨중의 하나,
$H_i(j)$: i번째 프레임의 j레벨에 대한 빈도

위 식에서 SD_i가 주어진 임계치보다 클 때, 장면의 변화가 일어났다고 간주한다. Nagasaka와 Tanaka[2]는 다음과 같은 식을 이용하여 장면변화를 추출한다.

$$SD_i = \sum_{j=1}^{G} \frac{| H_i(j) - H_{i+1}(j) |^2}{H_{i+1}(j)} \qquad (2)$$

이 방법은 히스토그램 비교 방법에서 두 프레임 사이의 차이를 강조하여 카메라 이동이나 물체 이동에 따른 미세한 작은 차이를 크게 한다.

이런 방법들은 선정된 임계치가 카메라 동작보다 낮을 때 오검출 가능성이 있으며, 너무 높으면 점진적 컷은 거의 검출하지 못한다. 본 논문에서는 이들 방법을 사용하며 임계치 설정 문제점을 해결하기 위해서 식(3)과 같은 방법을 이용하여 자동으로 설정한다. 이 방법은 이웃하는 두 히스토그램 차이의 절대값이 두 히스토그램 중 최소값 보다 크면 장면 변화가 발생했다고 정의한다. 따라서 이 방법은 프레임들 사이에 각기 다르

계 임계값이 설정된다. 또한 인덱싱 속도문제를 고려하여 모든 프레임(frame)을 비교하지 않고 스킵인자(skip factor)를 주어 일정한 간격으로 비교하는 방법을 이용한다. 스킵인자 설정은 검출의 주요 대상이 앵커 장면이므로 평균적인 앵커장면 길이의 1/3 값으로 설정한다. 앵커장면의 평균길이는 여러 실험 데이터를 사용하여 구한다.

$$|a_{i+1}-a_i| > \min(a_i, a_{i+1})\times b \quad b : 상수 \qquad (3)$$

$F_i \quad F_{i+1} \quad F_{i+2} \quad F_{i+3}$: *Frame*
↓ ↓ ↓ ↓ *(i=0 ~ end frame)*
$H_i \quad H_{i+1} \quad H_{i+2} \quad H_{i+3}$: *Histogram*
↘ ↙ ↘ ↙ ↘ ↙ *(i=0 ~ end frame)*
$a_i \quad a_{i+1} \quad a_{i+2}$: *Histogram difference.*
(i=0 ~ end frame-1)

2.2 뉴스사건의 분할과 아이콘 추출

디지털 뉴스 비디오에서 물체의 추출은 분할된 비디오의 내용을 이해하고 내용을 간단한 이미지로 표현하기 위해 사용한다. 본 논문에서는 뉴스가 앵커를 중심으로 구성되어 있는 성질을 이용하여 뉴스 사건을 분할한다. 즉, 뉴스의 내용을 함축적으로 포함하고 있는 뉴스 아이콘을 추출하기 위해 먼저 앵커영상을 검출하여 뉴스 아이템을 분할하고 다음으로 뉴스의 아이콘을 추출한다. 분리된 비디오에서 먼저 뉴스 앵커(anchor-person) 영상을 검출하기 위해서는 다음과 같은 정보를 이용한다. 국내 뉴스의 경우 앵커의 얼굴색상이 외국의 경우와 달리 단일 색상 분포를 가지고 있는 칼라 정보와 분할된 비디오의 첫 프레임들 중에 앵커의 장면이 존재한다는 정보를 이용한다.

본 논문에서는 이와 같은 정보를 사용하였고 앵커장면 검출은 영상에서 사람 영역을 추출하는 방법과 움직임이 거의 없다는 특징 및 앵커가 설명하고 있는 동안의 배경이 일정하다는 특징 그리고 앵커의 장면이 다른 분할된 프레임들보다 일반적으로 길게 나타난다는 특징을 이용한다. 첫 번째와 두 번째의 경우는 일반성이 있으나 시간이 많이 소요된다는 단점이 있다. 세 번째와 네 번째의 경우는 시간적으로 빠르다는 장점이 있는 반면에 장면 추출의 일반성이 없다는 단점이 있다. 따라서 이러한 문제점을 개선하기 위해서 다음절과 같은 방법을 사용한다.

2.2.1 앵커장면 검출에 의한 뉴스사건 분할

앵커의 영상을 추출하기 위해서는 미리 얼굴 영역의 색상 분포를 분석한다. 분석한 결과를 색상 공간에서 히스토그램으로 표현하며 사람간의 얼굴 색상 차이에 의한 문제를 줄이기 위해서 여러 뉴스에 들어있는 앵커의 얼굴영상을 이용하여 히스토그램으로 구성한다.

본 실험에서는 뉴스에 들어있는 앵커의 얼굴들을 실험영상으로 사용하여 색상분포를 구성하였으며 히스토그램 구성과정은 다음과 같다.

(그림 5) 실험 영상
(Fig. 5) Experimental image

먼저 (그림 5)와 같이 얼굴영상에서 얼굴부분의 영역을 오려낸다. RGB 컬러모델로 표현된 얼굴영역을 HSI 컬러모델로 변환한다. HSI 컬러모델은 색채요소(H,S)와 명암요소(I)를 분리하여 명암요소를 제거함으로써 조명변화의 영향을 줄일 수 있는 장점을 갖는다. RGB를 HSI로의 변환은 다음과 같은 식을 이용하여 변환한다[11, 13].

$$I = \frac{1}{3}(r+g+b)$$

$$S = 1-\frac{3}{(r+g+b)}[Min(r,g,b)] \qquad (4)$$

$$H = \cos^{-1}\left\{\frac{\frac{1}{2}[(r-g)+(r-b)]}{[r-g]^2+(r-b)(g-b)]^{\frac{1}{2}}}\right\}$$

히스토그램 구성을 위해서는 HSI 요소중 색상요소(H)와 채도 요소(S)만을 사용한다. 얼굴영역의 각 화소에 대해 H와 S를 구한 후 히스토그램의 해당 요소를 증가시킨다. H와 S는 히스토그램에서 좁은 영역에 집중된다. 구성된 히스토그램을 이용하여 장면이 분할된 뒤 각 화소의 컬러 값이 얼굴 색상 가능성 범위에 들어오면 앵커영상의 후보로 인식된다.

이와 같이 구해진 얼굴 색상 정보를 이용하여 뉴스에서 앵커가 처음에 등장하므로 처음 컷에서 사람영역을 인식하여 앵커의 배경색상과 모양성보를 획득한다. 후보로 등록된 영상은 다음 기준에 따라 앵커영상의 검

증과정을 거친다. 검증과정은 에러를 최대로 줄이기 전역적인 비교법과 지역적 비교법 그리고 움직임 벡터를 채택한다.

• 앵커영상의 공간적 배경이 비슷하다는 정보를 이용하여 배경의 색상분포에 대한 검증

• 앵커영상은 영상의 움직임이 거의 발생하지 않은 정보를 이용하여 움직임 벡터(motion vector)계산에 의한 검증

첫 번째로 앵커의 배경영상 검증은 앵커 영상의 공간적 구조가 뉴스 아이콘을 제외하고는 항상 색상분포가 동일하다는 특징을 이용한다. 그 방법으로는 전체 영역의 색상분포와 (그림 6)과 같이 다른 물체가 나타나지 않는 부분 영역의 색상분포 각각에 히스토그램 교차 방법을 적용하여 검증한다. 본 논문에서는 색상 히스토그램에 기반한 두 영상의 유사성을 측정하기 위해 Swain과 Ballard의 히스토그램 교차 방법[A. Nagasaka 92]를 적용한다[2]. 영상의 색상 히스토그램은 이산 영상 칼라(bins)라고 불리는 RGB로 분류하고 영상의 모든 픽셀은 교차에 의해 나타나는 이산 칼라의 수를 카운트함으로써 얻어진다. 두 이미지 f_i와 f_j, 그리고 두 이미지의 히스토그램 I_i(참조 화상)와 I_j(질의 화상)가 주어졌을 때 각각은 n bins로 구성한다. 여기서 k번째 빈에서 I_i의 값을 $I^i{}_k$로 표기한다. 히스토그램의 교차는 식 5와 같이 정의한다.

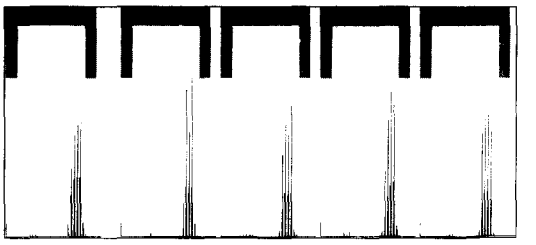
$$\sum_{k=1}^{n} \min(I^i{}_k, I^j{}_k) \qquad (5)$$

식 5는 두 영상에서 유사한 색상들을 가진 픽셀의 수를 인지한다. 또 다음과 같이 0과 1사이의 값으로 매칭하기 위해 식 (6)과 같이 일반화하여 사용한다.

$$S(I^i, I^j) = \frac{\sum_{k=1}^{n} \min(I^i{}_k, I^j{}_k)}{\sum_{k=1}^{n} I^j{}_k} \qquad (6)$$

(그림 6)에서 첫 번째 영상은 뉴스의 처음 앵커 영상의 배경 히스토그램이며 나머지 영상은 위의 방법에 의하여 검출된 영상이다.

두 번째로 움직임의 검증은 움직임 벡터를 계산한다. 카메라와 물체의 상대적 움직임에 의해 발생된 움직임 벡터는 연속하는 두 프레임 사이의 움직임을 추정하는 단위에 따라 크게 화소 순환법(PRA)과 블럭 매칭법(BMA)으로 나뉜다. 블록 매칭법은 블럭 단위로 움직임을 추정하기 때문에 순 정확도는 떨어지지만 알고리즘이 단순하여 구현에 용이하다는 이점이 있다. 따라서

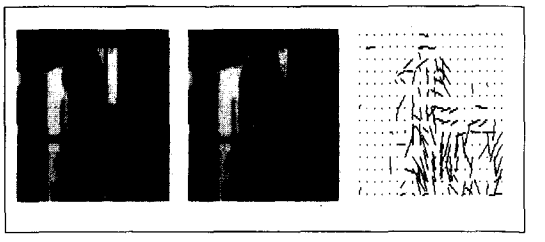


(그림 6) 배경색상의 히스토그램
(Fig. 6) Histogram of background color

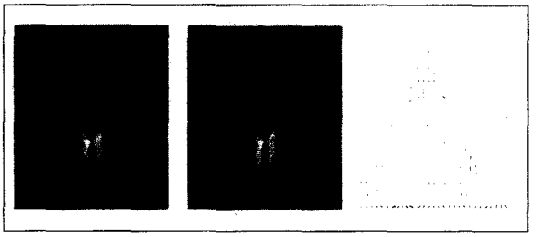
본 논문에서는 현재 프레임을 일정한 크기의 블럭으로 나누고 시간축상 기준이 되는 프레임(reference frame)에서 각 블럭의 움직임을 추정하는 방식으로 보편적인 블록 매칭법을 사용하며 블럭의 유사도를 측정하는 평가 함수로 식 (7)에서 정의된 평균제곱오차(MSE:Mean Square Error)를 사용한다.[1,9]

$$MSE = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} [S_k(m,n) - S_{k+1}(m,n)]^2 \qquad (7)$$

여기서 Sk와 Sk+1은 각각 k번째와 k+1번째 프레임이다. 일반적으로 탐색 영역내의 모든 후보 변위 벡터에 대해서 기대 오차를 계산한 후 가장 작은 오차를 갖는 후보 변위 벡터를 현재 블럭의 움직임 벡터로 채택한다. 움직임 벡터는 방향과 크기로 이루어지는데, 방향성분은 카메라의 동작(패닝, 틸팅, 줌)을 검출하기 위함이고, 크기 성분은 카메라 동작의 속도를 나타내기 위함이다. 본 논문에서는 이들 방법을 사용하여 방향성분과 크기 성분을 각각 계산한 후 크기가 임계치 이상인 경우에 8방향(8-neighbourhood)별로 누적한다. 그리고 이들 분포값이 작은 경우에 앵커영상으로 판단한다(그림 7)(그림 8). 이와 같이 배경정보와 움직임 벡터 정보를 이용하여 검증된 앵커영상은 새로운 뉴스의 시작으로 판단하고 뉴스 아이템으로 분할하여 인덱싱 정보로 사용한다.



(그림 7) 움직임 있는 모션벡터
(Fig. 7) Motion vector of motion object



(그림 8) 움직임이 없는 모션벡터
(Fig. 8) Motion vector of motionless object

2.2.2 뉴스 아이콘 추출

검출된 앵커 영상은 앵커 위치에 따라 세 가지 모델로 분류되며 뉴스 아이콘의 존재 여부가 결정된다. 즉, 앵커의 위치가 우측이면 아이콘은 좌측 상단에 존재하고, 앵커의 위치가 좌측이면 아이콘은 우측 상단에 존재한다. 앵커의 위치가 중앙에 있을 때는 뉴스 아이콘은 앵커 장면에 존재하지 않는다〈표 1〉.

앵커의 위치 결정은 검출된 앵커얼굴의 칼라 분포로서 다음과 같은 방법에 따라 결정한다. 먼저 세 가지 모델에 따라 각각의 소속정도를 계산한 후 식(8)과 같이 최대 값이 임계치(θ)보다 충분히 크면 분류한다. 최대 값이 유일하면 최대 값을 그 모델로 분류하고 그 외의 경우는 분류하지 않는다.

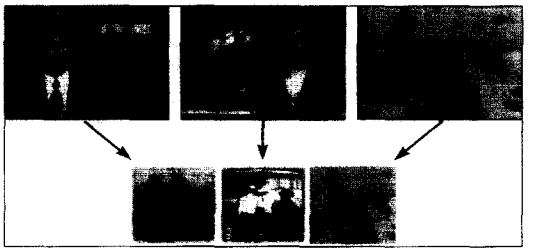
$$\text{Max}\{AL, AC, AR\} > \theta \quad (8)$$

θ : 미리 계산된 얼굴면적의 최소 값

이와 같은 방법에 따라 모델이 분류되면 뉴스아이콘을 추출한다. 뉴스 아이콘은 뉴스에서 항상 같은 위치에 생성되므로 미리 모델에 따라 아이콘의 위치를 계산한 후 추출한다. 뉴스아이콘이 존재하지 않은 모델에서는 뉴스의 시작 컷에서 대표 프레임을 추출하여 뉴스 아이콘의 크기로 축소하여 뉴스 아이콘을 생성한다. 생성된 아이콘은 뉴스를 대표하는 부분 영상으로서 뉴스 인덱싱 정보로 사용한다. (그림 9)는 추출된 뉴스 아이콘의 정보를 나타낸다.

〈표 1〉 뉴스 아이콘의 정보
〈Table 1〉 Information of news icon

앵커위치 / 아이콘	좌측 (AL)	우측 (AR)	중앙 (AC)
아이콘 존재여부	존재함	존재함	존재 않음
아이콘 추출	우측 상단	좌측 상단	앵커 다음컷

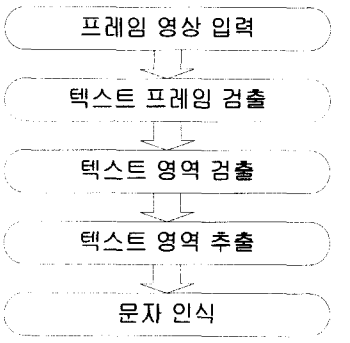


(그림 9) 추출된 뉴스 아이콘
(Fig. 9) Extracted news icon

2.3 자막 인식

비디오 영상에 나타날 수 있는 텍스트의 종류로는 먼저 비디오 영상 내에 포함되어 있는 텍스트로서 글자의 크기와 형태 등이 일정하지 않은 씬 문자(scene text)가 있으며 이러한 형태의 텍스트는 추출하기가 어렵다. 다음으로 비디오 영상에 특정한 내용 또는 설명을 부여하기 위하여 인위적으로 비디오 영상에 삽입된 텍스트로서 인위적인 문자(artificial text)가 있다. 뉴스의 자막은 인위적 텍스트의 대표적인 사례이며 이러한 텍스트의 특징으로는 문자가 전경(foreground) 형태이고 단색이며, 문자의 형식과 크기가 일정하다. 또한, 여러 프레임 사이에 지속적으로 표현되며, 제한된 영역에서 정지한 상태이거나 선형으로 움직인다. 그리고 읽기 쉽게 배경(background) 처리를 포함하고 있기도 한다[12].

본 논문에서는 이러한 인위적인 텍스트의 정보를 이용하여 텍스트를 추출하고 인식하여 인덱싱 구성 요소로 사용한다. 자막인식 과정은 텍스트 프레임 검출, 문자열 추출, 그리고 문자 인식의 세 단계로 구성된다(그림 10).



(그림 10) 문자열 추출과정
(Fig. 10) String extraction process

2.3.1 텍스트 프레임 검출과 문자열 추출

텍스트 프레임 검출은 자막 인식을 위한 첫 번째 단계로 입력되는 프레임이 텍스트 프레임인가를 판별하는 과정이다. 검출 과정은 먼저 자막 배경색상에 대한 히스토그램을 구한다. 다음으로 텍스트의 전경색상 특징으로 텍스트가 배경과 대조적이고 텍스트에 윤곽선 또는 경계를 갖는 특징을 이용하여 히스토그램을 구한다. 이 과정은 식(9)과 (10) 같이 입력 영상에 대하여 y방향으로 이웃한 픽셀들의 색상 차이에 의하여 판별한다.

$$Pd(x,y)=\begin{cases}1, & |F(x,y)-F(x+1,y)| > t \\ 0, & otherwise\end{cases} \quad (9)$$

여기에서 F는 프레임이고, (x, y)는 픽셀의 위치이다. t는 이웃한 픽셀의 색상 차이에 대하여 설정한 임계치이다.

$$Pdy(x)=\sum_{y=1}^{Y-1} Pd(x,y) > T \quad (10)$$

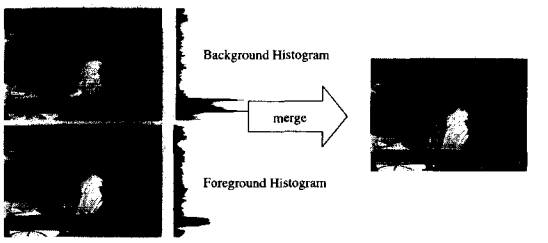
식(10)에서 $Pdy(x)$는 y축 방향으로 픽셀의 색상 차이가 t이상 변한 개수의 합이다. Y는 y축 해상도이고, T는 임계치이다.

마지막으로 계산된 두 히스토그램을 상호 병합하여 임계치 이상인 경우에 텍스트 프레임 후보로 검출한다. 검출된 프레임은 뉴스의 자막이 일정시간 계속된다는 특성을 이용하여 연속하는 세 프레임 이상에서 텍스트 프레임 검출과정을 반복하여 똑같은 결과일 때 텍스트가 존재하는 프레임으로 간주한다.

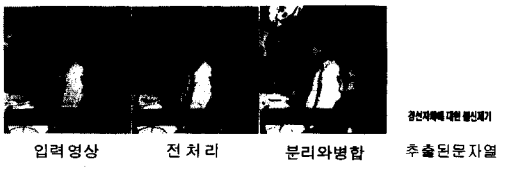
텍스트 영역 검출은 검출된 텍스트 프레임에서 인위적인 텍스트 조건에 의하여 선별하는 과정이며 또한 텍스트 추출에서 영역 정보를 이용할 수 있다. 텍스트 프레임 검출에서 계산된 $Pdy(x)$가 임계값에 만족하는 연속적인 구간이 너무 작거나 또는 너무 크면 인위적인 텍스트 특징에 위반되므로 이러한 구간은 영역 검출에서 제외된다. 영역 검출 과정은 입력 영상에 대하여 식(9)과 (10)를 이용하여 x 방향으로 이웃한 픽셀들의 색상 차이에 의하여 판별한다. 텍스트 영역 검출은 텍스트 프레임 추출과정에서 계산된 y축 데이터와 x축 데이터를 결합하여 텍스트 영역을 결정한다. (그림 11)는 이와 같은 과정을 보여준다.

검출된 텍스트 프레임은 전처리 후에 Horowitz와 Pavlidis가 제안한 Split-and-Merge 알고리즘을 이용하여 문자열을 추출한다[13]. 문자열 추출과정은 다음과 같은 순서를 따른다. 첫 번째로 전처리 과정은 문자열의 컬러 색상대비(contrast)를 높이는 과정을 거친다. 이렇게 함으로써 영역내에 들어있는 컬러의 수를 줄일 수 있다. 또한 문자 추출시 문자가 단절되는 것을 없애기 위해서 문자의 색상에 해당되는 픽셀을 한 픽셀씩 확장한다.

다음으로 분할과 병합(split-and-merge)방법을 이용하여 문자열을 추출한다. 분할(split) 단계에서는 전체 영상을 다른 크기의 정방형(square)영역들로 분할하고, 병합(merge) 단계에서는 인접한 정방형 영역들 사이에 동질성(homogeneity)을 조사하여 비정방형 영역으로 합병을 해 나간다. 마지막으로 문자의 후보영역으로 적당하지 않은 부분을 제거한다. 이 과정에서는 문자의 폭과 높이가 임계치 보다 너무 크거나 작은 경우에는 문자가 아닌 것으로 판단한다. (그림 12)는 문자열 추출과정을 보인다.



(그림 47) 텍스트 프레임 검출
(Fig. 11) Text frame detection

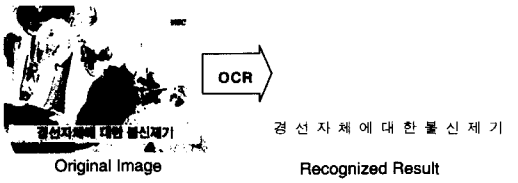


(그림 48) 문자열 추출과정
(Fig. 12) Process of characters extraction

2.3.2 문자 인식 및 검색

추출된 문자열 영상은 사용자가 질의를 사용하여 검색할 수 있도록 이미 개발된 문자 인식기(OCR)에 입력하여 결과를 얻는다. 이러한 인식기의 문자인식은 한글과 영문을 자소별로 분할하여 인식하는 방법을 사용한다. (그림 13)는 추출된 문자열 영역에서 인식된 결과를 나타낸다. 하지만 이러한 인식기는 정지영상에 기반한 인식기로써 비디오 영상에는 인식률이 좋지 않다. 따라서 검색의 효율을 높이는 방안으로 인식된 문자열

로 인덱싱하여 검색하는 방법과 추출된 텍스트 영역을 인덱싱하여 검색하는 방법 모두를 이용하였다. 향후 비디오 영상에 기반한 인식기의 개발에 관한 연구도 있어야할 것으로 생각된다. 인식된 문자열은 (그림 4)과 같이 인덱스 파일의 구성 정보로 사용한다. 인덱스 파일 구조는 인식된 문자열과 프레임 시작번호와 종료번호로 구성하며 사용자의 질의에 의하여 원하는 프레임(frame)을 쉽게 검색할 수 있도록 하였다.



(그림 49) 인식된 문자열
(Fig. 13) Recognized characters

3. 검색 브라우저

일반적으로 동영상 데이터와 같은 비정형 데이터를 검색하기 위해서는 동영상 자체를 처리하여 의미있는 단순한 이미지들의 집합으로 나타내는 방법이나 이미지와 연관된 텍스트를 분석하여 검색하는 방법을 이용한다. 기존의 이미지 검색시스템에서는 이미지 데이터들을 일일이 보면서 검색하거나 이미지 데이터에 대한 속성정보 검색만을 지원했다. 그러나 이미지 데이터의 특성을 고려해 볼 때 이미지 데이터를 일일이 보면서 원하는 이미지를 찾아간다는 것은 거의 불가능하다. 왜냐하면 이미지 데이터는 일반적으로 데이터 크기가 크면 데이터를 읽어오는 속도가 느려 실시간에 보여 주면서 검색하는데 한계가 있기 때문이다[3, 6].

본 논문에서는 첫 번째, 아이콘 기반 검색 방법으로는 (그림 3)과 같은 인덱스 파일 정보를 이용하여 비디오의 순차적인 줄거리를 가시화하는 방법으로, 장면전환 영상이나 추출된 내용을 아이콘화하여 검색할 수 있도록 하였다. 또한 뉴스 사건별로 구분하여 원하는 동영상 정보를 빨리 검색할 수 있도록 설계하였다. 사용자 인터페이스의 전체화면은 뉴스 아이콘 윈도우, 자막문자 아이콘 윈도우, 뉴스 플레이 윈도우 세 부분으로 구성하였다.

두 번째, 내용기반 검색 방법으로는 인식된 문자열을 사용하여 사용자의 질의가 가능하도록 하였다. 사용자 인터페이스는 질의창(query window)과 검색 결과창(query result window)으로 구하였다. 사용자의 질의에 의한 검색은 불리안 식을 기반으로 정확한 스트링 매칭과 근사한 스트링 매칭 두 가지 검색 방법이 가능하도록 하였다.

4. 실험 및 결과

본 논문의 구현은 Windows-NT 환경에서 Visual C++ 2.0의 API 함수를 사용하여 구현하였다. 실험 데이터는 현재 방송되고 있는 국내의 MBC 방송국 뉴스를 압축된 AVI(640×480, 24 프레임/초) 파일 형식으로 실험하였다.

〈표 2〉에서 열 방향은 실험 데이터의 순서이며 행 방향은 장면전환과 뉴스사건 그리고 자막문자 검출 결과이다. N_t는 데이터에 존재하는 실제의 개수, N_d는 검출된 개수이고 N_e는 오검출의 개수를 나타낸다. 검출율(P)은 식(11)에 의하여 평가된다.

$$\text{검출율}(P) = \frac{N_d}{N_t} \times 100 \tag{10}$$

〈표 2〉 실험 결과
〈Table 2〉 Experimental result

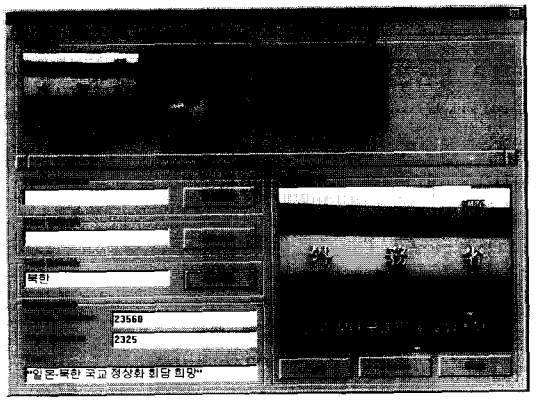
검출 / 뉴스	장면전환 검출				뉴스사건 검출				자막문자 검출			
	Nt	*Nd*	*Ne*	*P*	*Nt*	*Nd*	*Ne*	*P*	*Nt*	*Nd*	*Ne*	*P*
MBC뉴스 #1	267	247	20	92.5	12	11	1	91.7	15	14	1	93.4
MBC뉴스 #2	118	102	16	86.5	7	6	2	85.7	7	7	0	100
MBC뉴스 #3	111	103	8	92.8	13	15	2	84.7	10	9	1	90

(그림 14)는 실험에 사용한 데이터를 나타내고 있다. (그림 15)는 뉴스아이콘과 자막정보를 이용하여 뉴스 사건별로 검색할 수 있는 아이콘기반 검색 브라우저를 나타내고 있으며, (그림 16)는 내용에 기반하여 질의가 가능한 내용기반 검색 브라우저를 나타내고 있다.

5. 결론 및 향후 연구 방향

본 논문은 구조화된 뉴스 비디오의 자동 인덱싱과

(그림 14) 실험 데이터
(Fig. 14) Experimental data



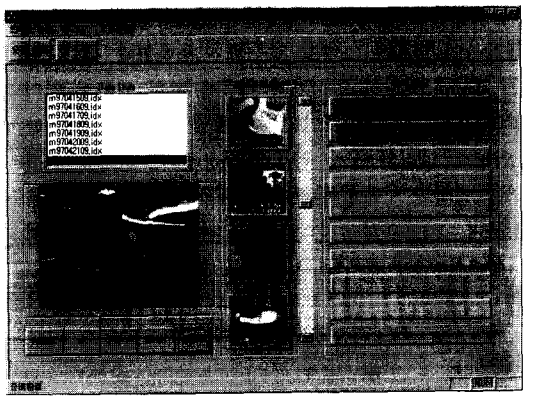
(그림 16) 내용기반 검색 브라우저
(Fig. 16) Content-based retrieval browser

검색을 위한 통합된 방법을 제시하였다.

인덱싱 방법은 뉴스사건의 분할에 의한 뉴스아이콘 추출과정과 자막문자 추출에 의한 자막인식과정으로 이루어져 있다. 이러한 인덱싱 방법은 뉴스 비디오에 대하여 아이콘기반 검색과 내용기반 검색이 가능함을 증명하고 있다. 구현된 검색 브라우저는 시간을 소비하는 기존의 순차적인 뉴스 비디오 검색 방법의 문제점을 해결하도록 설계되고 구현되었다.

실험결과는 현재 방송되고 있는 국내의 MBC 뉴스를 대상으로 하였으며 아주 효과적으로 인덱싱됨을 보이고 있다. 따라서 본 연구에 기반하여 다른 도메인에 쉽게 확장할 수 있을 것이며, 또한 주문형 뉴스(NOD)를 제공하는데 효율적으로 적용될 것으로 기대된다.

향후, 효율적인 비디오 검색을 위해서는 음성 인식에



(그림 15) 아이콘기반 검색 브라우저
(Fig. 15) Icon-based retrieval browser

의한 인덱싱 방법이 필요하며, 여러 가지 모델에 적용하여 신뢰성 있는 모델로 개선해야 할 것으로 생각된다.

참 고 문 헌

[1] A. Akutsu et al., "Video Indexing Using Motion Vector," Proc. SPIE Visual Comm. and Image Processing 92, Bellingham, Wash., Vol. 1818, pp. 1522-1530, 1992.

[2] A. Nagasaka and Y. Tanaka, "Automatic Video Indexing and Full-Video Search for Object Appearances," Visual Database System, Vol. 11, E. Knuth and L.M.Wegner, eds., Elsevier, Amsterdam, pp. 113-127, 1992.

[3] A. Yoshitaka, S. kishida, M. Hirakawa,and T. Ichikawa, "Knowledge-Assisted Content-Based Retrival for Multimedia Database," IEEE Multimedia, Winter pp. 12-21, 1994.

[4] H. Ueda et al., "Automatic Structure Visualization for Video Editing," Proc. InterCHI 93 ACM Press, New York, pp. 137-141, 1993.

[5] H. Zang, A. Kankanhalli, and S. Smoliar, "Automatic Partitioning of Full-Motion Video", Proc. ACM Multimadia System, New York, Vol. 1, pp. 10-28, 1993.

[6] H.J. Zhang and S.W. Smoliar, "Developing Power Tools for Video Indexing and Retrival," Proc. IS&T/SPIE Symp. Electronic Imaging: Science and Technology, 1994.

[7] H.J. Zhang et al., "Automatic parsing of News Video," Proc. IEEE Int'l Conf. Multimedia Computing and System, IEEE Computer Society Press, Los Alamitos Calif., 1994.

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

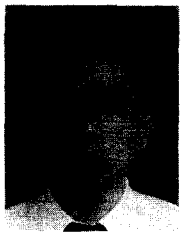
[9] HongJing Zhang, Shuang Yeo Tan, Stephen W. Smoliar, Gong Yihong, "Automatic parsing and indexing of news video," Multimedia Systems, 1996.

[10] K. Otsuji and Y. Tonomura, "Projection Detecting Filter for Video Cut Detection," Proc. ACM Multimadia 93, ACM Press, New York, pp. 251-257, 1993.

[11] M. Miyahara and Y. Yoshida, "Mathematical Transform of (R, G, B) Color Data to Munsell(H, V, C) Color Data," Proc. SPIE Visual Comm. and Image Processing 88, SPIE, Bellingham, Wash., Vol. 1001, pp. 650-675, 1988.

[12] Rainer Lienhart, "Automatic Text Recognition for Video Indexing," Proc. ACM Multimedia 96, Boston MA USA, pp. 11-20, 1996.

[13] S. L. Horowitz and T. Pavlidis, "Picture Segmentation by a Traversal Algorithm," Comput. Graphics Image Process. 1, pp. 360-372, 1972.



양 명 섭

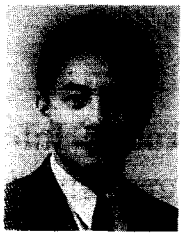
1990년 전북대학교 이학계열 졸업(이학사)

1995년 전북대학교 대학원 전산통계학과 졸업(이학석사)

1997년 전북대학교 대학원 전자계산학과 박사과정 수료

1997년 ~현재 전북대학교 대학원 전자계산학과 박사과정

관심분야 : 비디오정보검색, HCI, 분산처리, 컴퓨터그래픽스 등



유 철 중

1982년 전북대학교 전산통계학과 졸업(이학사)

1985년 전남대학교 대학원 계산통계학과 졸업(이학석사)

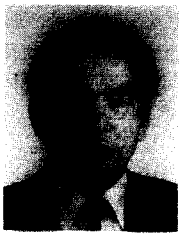
1994년 전북대학교 대학원 전자계산학과 졸업(이학박사)

1982년~85년 전북대학교 전자계산소 조교

1985년~96년 전주기전여자대학 전자계산과 근무

1997년~현재 전북대학교 자연과학대학 컴퓨터과학과 전임강사

관심분야 : 소프트웨어공학, 객체지향시스템, 멀티미디어, HCI, 분산객체컴퓨팅 등



장 옥 배

1966년 고려대학교 수학과 졸업(학사)

1974년~80년 죠지아주립대, 오하이오주립대 박사과정 수료

1990년~91년 영국 에딘버러대학교 객원교수

1980년 ~현재 전북대학교 컴퓨터과학과 교수

관심분야 : 소프트웨어공학, 전산교육, 수치해석, 인공지능 등