

논문 2004-41CI-3-12

# 한국어 폐쇄자막을 이용한 지식기반 비디오 검색 시스템

## (Knowledge-based Video Retrieval System Using Korean Closed-caption)

조 정 원\*, 정 승 도\*\*, 최 병 욱\*\*\*

(Jung-Won Cho, Seung-Do Jeong, and Byung-Uk Choi)

### 요 약

저 수준의 특징정보를 사용하는 내용기반 검색만으로 지능형 정보검색을 위한 사용자의 개념적인 요구에 부합하는 검색결과를 제공하기 어렵다. 일반적으로 비디오 데이터에는 동영상 정보와 함께 음성, 음향 등의 오디오 정보와 폐쇄자막 등의 정보가 포함되어 있다. 지식기반 비디오 검색은 그러한 다양한 정보를 사용하여 자동색인을 수행하고 색인 데이터베이스를 구축한다. 이로써 사용자는 보다 개념적인 검색 요구에 부합하는 검색 결과를 얻을 수 있다. 본 논문에서는 비디오 내의 한국어 폐쇄자막을 이용한 지식기반 비디오 검색 시스템을 제안한다. 한국어 폐쇄자막은 형태소 분석 수준에서 자동색인되며, 색인 데이터베이스를 이용하여 키워드 질의를 통해 비디오를 검색할 수 있다. 실험에서 한국어 속기시스템으로 제작된 폐쇄자막이 포함된 뉴스비디오에 적용하여, 제안하는 방법이 사용자의 보다 의미 있는 개념적인 요구에 부합하는 검색 결과를 얻을 수 있음을 확인하였다.

### Abstract

The content-based retrieval using low-level features can hardly provide the retrieval result that corresponds with conceptual demand of user for intelligent retrieval. Video includes not only moving picture data, but also audio or closed-caption data. Knowledge-based video retrieval is able to provide the retrieval result that corresponds with conceptual demand of user because of performing automatic indexing with such a variety data. In this paper, we present the knowledge-based video retrieval system using Korean closed-caption. The closed-caption is indexed by Korean keyword extraction system including the morphological analysis process. As a result, we are able to retrieve the video by using keyword from the indexing database. In the experiment, we have applied the proposed method to news video with closed-caption generated by Korean stenographic system, and have empirically confirmed that the proposed method provides the retrieval result that corresponds with more meaningful conceptual demand of user.

**Keywords :** Automatic indexing, Closed-caption, Time-alignment, Video retrieval

### I. 서 론

디지털 멀티미디어 데이터는 나날이 증가하고 있다. 이 방대한 정보에 대한 효율적인 관리 및 이용은 필수적이다. 정보검색 시스템은 그러한 데이터를 수집하여

특징정보를 추출하고 저장한 후, 이를 이용하여 사용자가 원하는 데이터를 찾아주는 시스템이다. 이 시스템에서의 주요 관심은 사용자의 요구에 적합한 데이터를 빠르게 찾아주는데 있다. 인터넷의 대중화와 초고속 정보통신망의 구축은 정보교류의 수단을 텍스트에서 이미지, 그리고 비디오에 까지 이르게 하였다. 그림 1은 주요 검색대상의 변화에 따른 검색기술의 변천을 나타내고 있다. 정보검색 기술은 텍스트 처리를 통한 키워드 기반 텍스트 검색에서 이미지의 색상(color) 및 형태(shape), 질감(texture) 등의 저 수준 특징정보(low-level feature)를 추출하는 내용기반 이미지 검색

\* 정회원, \*\* 학생회원 한양대학교 전자통신전공학과 (Department of Electrical and Computer Engineering, Hanyang University)

\*\*\* 정회원, 한양대학교 정보통신대학 정보통신학부 (Division of Information and Communications, Hanyang University)

접수일자: 2003년10월14일, 수정완료일: 2004년04월20일

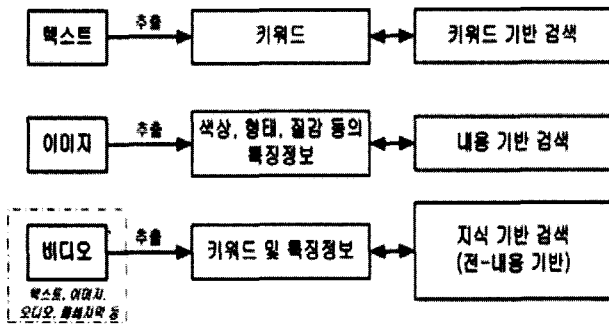


그림 1. 검색대상에 따른 검색방법  
Fig. 1. Retrieval methods according to the main subject for information retrieval.

(content-based image retrieval)<sup>[1]</sup>으로 변화되었다. 지식기반 비디오 검색<sup>[2]</sup>은 비디오에 포함된 동영상 및 오디오 정보, 폐쇄자막(closed-caption) 등의 모든 콘텐츠를 이용하는 것으로 전-내용(full-content)기반 검색이라고도 한다. 이 지식기반 검색은 사용자의 개념적인 요구에 보다 적합한 결과를 얻을 수 있다.

본 논문에서는 지식기반 비디오 검색을 위한 한국어 폐쇄자막의 자동색인 방법과 색인된 데이터베이스를 이용한 검색방법을 제안한다. 폐쇄자막을 이용하기 위해서는 비디오 내의 동영상 및 오디오 정보와 동기된 시간을 할당하는 것이 매우 중요하다. 폐쇄자막의 시간할당을 각 어절 단위로 자막디코더에 수신되는 시간을 직접 기록하여 처리하여, 잘못 동기되는 오류를 최소화할 수 있다. 이 결과만으로도 폐쇄자막에 포함된 키워드에 대한 비디오 검색이 가능한 기본검색 서비스를 제공할 수 있다. 사용자의 개념적인 요구에 보다 부합하는 검색 결과를 얻기 위해서 폐쇄자막에 대한 형태소분석 및 분석결과의 모호성해소를 수행하고, 미등록어 처리, 복합명사 처리, 불용어 제거의 과정을 거쳐 색인어를 선정한다. 고급검색에서는 그 색인 데이터베이스를 이용하여 상세검색을 지원하며, MPEG-7의 칼라구조 기술을 이용한 내용기반 이미지 검색도 지원한다. 실험에서 한국에서 공통적으로 적용되고 있는 속기시스템이 생성한 폐쇄자막이 들어 있는 한국 뉴스 비디오에 적용한다. 실험의 결과로, 제안하는 방법이 사용자의 개념적인 요구를 반영할 수 있는 비디오에 대한 복합적인 질의가 가능하기 때문에 그 요구에 보다 부합하는 결과를 제공할 수 있음을 확인하고자한다.

본 논문의 구성은 다음과 같다. II장에서는 자막의 개념과 종류에 대해 기술하고, III장에서는 폐쇄자막을 검색의 주요 정보로 사용하기 위한 시간할당 방법에 대



그림 2. 자막의 종류  
Fig. 2. Captions in video.

해 기술한다. IV장에서는 자동색인 과정으로 비디오 분할, 폐쇄자막의 키워드추출, 프레임에서의 특징정보 추출방법에 대해 설명하며, V장에서 기본검색과 상세검색 등의 검색방법에 대해 기술하고, VI장에서 폐쇄자막의 시간할당 및 비디오 분할 성능을 평가한다. 끝으로 VII장에서 결론을 맺는다.

## II. 자막의 개념과 종류

비디오는 동영상 및 오디오, 폐쇄자막 등의 데이터를 갖는다. 폐쇄자막은 비디오의 등장인물의 음성 내용뿐만 아니라 각 화자의 구분 및 화면 내에서의 화자의 위치, 주변 상황에 대한 설명 등이 포함되어 있어 검색에 유용하게 사용될 수 있다. 비디오에 포함된 자막은 크게 공개자막(open-caption)과 폐쇄자막 두 종류로 나뉜다. 폐쇄자막은 전용 디코더를 거쳐야만 화면으로 볼 수 있는 반면에, 공개자막은 디코더 없이도 화면상에서 볼 수 있는 문자들을 일컫는다. 그림 2는 TV화면에서 확인할 수 있는 공개자막과 폐쇄자막을 각각 보이고 있다. 폐쇄자막은 청각장애인을 위한 서비스로 시작되었다. 한국에서는 MBC가 1999년 2월에 처음으로 폐쇄자막이 포함된 방송을 시작하였으며, KBS, SBS가 뒤를 이었다. 폐쇄자막은 비디오에 포함된 음성정보를 그대

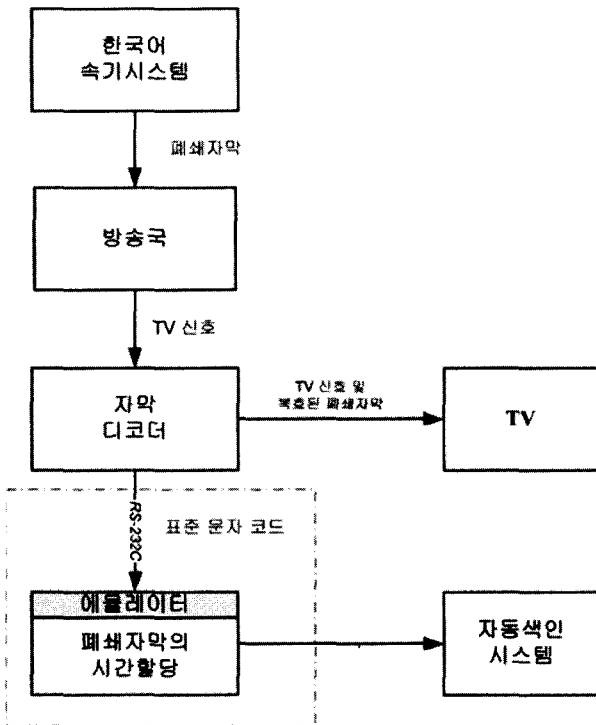


그림 3. 이진폐쇄 자막 추출 및 시간할당  
Fig. 3. Schematic diagram of extraction and time-alignment of the closed-caption.

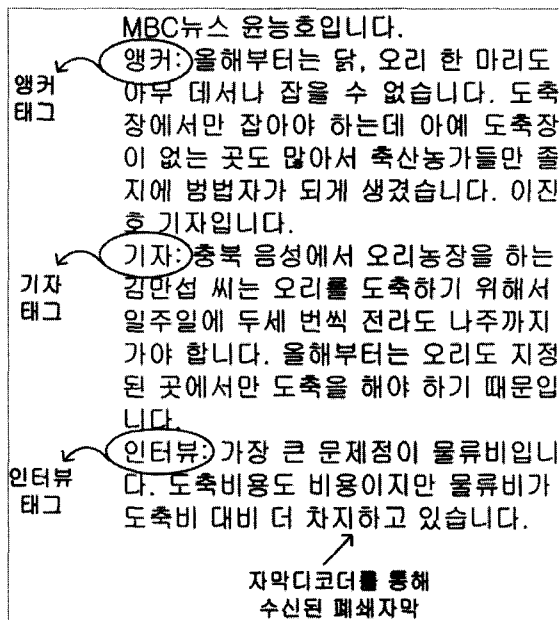


그림 4. 폐쇄자막에 포함된 화자 구분을 위한 태그  
Fig. 4. Tags in the caption.

로 나타내고 있어 비디오의 의미를 표현하는데 가장 적합한 데이터이다. 그림 3과 같이 별도의 인식 과정 없이 속기시스템에서 생성한 표준 문자 코드로 이루어진 폐쇄자막을 디코더를 통해 바로 얻을 수 있다. 이 폐쇄자막은 바로 동영상 및 오디오와 동기시킨 후, 기본검

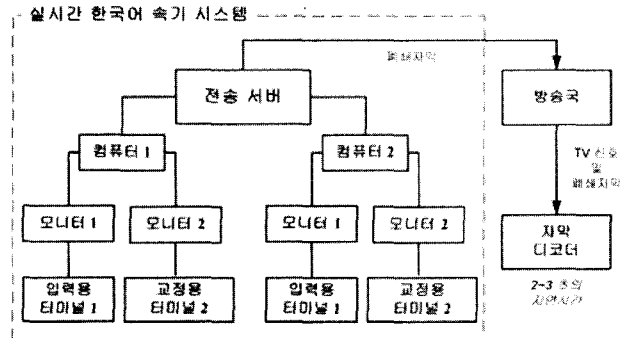


그림 5. 한국어 속기시스템의 구성  
Fig. 5. Schematic diagram of Korean stenographic system.

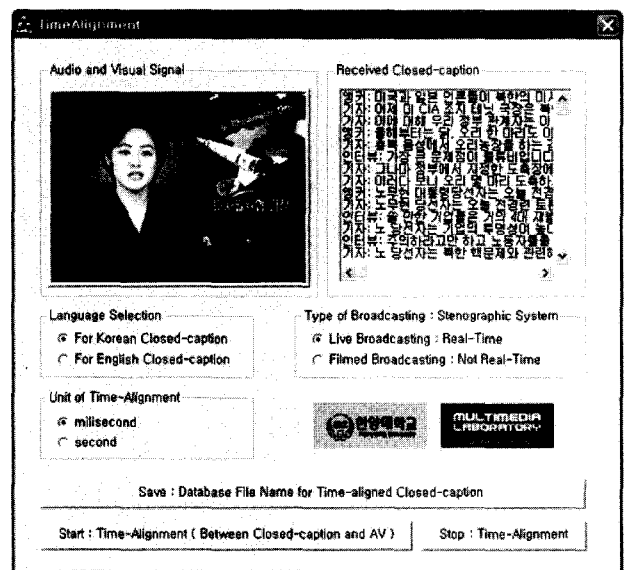


그림 6. 폐쇄자막의 시간할당 프로그램의 인터페이스  
Fig. 6. User interface of the time-alignment system.

색(basic search)에 바로 사용되거나 고급검색(advance search)을 위한 형태소분석 수준의 키워드 추출 시스템에 적용된다. 본 논문에서 실험대상으로 적용하고 있는 뉴스 비디오의 경우, 그림 4에서 보이는 바와 같이, 폐쇄자막에 앵커와 기자 등으로 이미 태깅되어 있어 비디오의 분할 작업에 효과적으로 사용될 수 있는 장점을 갖는다. 또한 영화나 드라마에 포함된 폐쇄자막의 경우에는 화자의 음성정보 외에도 영상과 음향에 대한 묘사 정보도 갖고 있어, 이 또한 검색에 효과적으로 사용될 수 있다.

### III. 폐쇄자막의 시간할당

폐쇄자막을 비디오에 포함된 동영상 및 오디오 정보와 동기된 시간으로 할당하는 것은 검색의 응용에 있어

count	word	time
734	앵커:	[379,21]
735	다음	[379,49]
736	소식입니다.	[380,89]
737	남이	[381,05]
738	들어	[381,29]
739	있는	[381,59]
740	꽃게가	[382,16]
741	오늘도	[382,49]
742	부산에서	[382,86]
743	발견됐습니다.	[383,93]
744	꽃게에	[384,09]
745	남을	[384,36]
746	남은	[384,66]
747	꽃이	[384,96]
748	중국의	[385,30]
749	머리	[385,76]
750	지역인	[386,50]
751	것으로	[386,83]
752	추정됩니다.	[387,86]
753	한편	[388,02]
754	지난	[388,73]
755	6월부터	[389,13]
756	남을	[389,70]
757	남은꽃게가	[390,34]
758	발견됐다는	[391,13]
759	머리가	[391,47]
760	나오고	[391,63]
761	있습니다.	[392,67]
762	부산의	[392,83]
763	비번호	[393,13]
764	기자입니다.	[395,90]

그림 7. 시간할당 후 데이터베이스에 저장된 폐쇄자막  
Fig. 7. Closed-caption in database after time-alignment.

필수적이다. 폐쇄자막을 이용하는 *Informedia* 프로젝트<sup>[2],[3]</sup>에서는 음성인식 과정을 통해 시간할당을 수행하였다. 이 방법도 성공적으로 폐쇄자막의 시간할당을 수행할 수 있으나, 음성인식 기술의 적용으로 인해 처리 시간이 길어지고, 음성이 녹음된 환경에 따라 성능이 영향을 받는 문제점이 있다. 특히, 뉴스 비디오에서는 스튜디오 밖에서 녹음된 기자의 음성에 잡음이 많이 포함되어 있고, 배경에 해당하는 사람들의 잡음이 포함되어 있는 경우도 빈번해 음성인식을 사용한 시간할당으로는 동기오류(false synchronization error)의 해결이 어렵다.

본 논문에서는 보다 정확하고 빠른 폐쇄자막의 시간할당을 위해 폐쇄자막을 생성하는 속기시스템을 분석하고, 이에 따른 시간할당 방법을 제안한다.

1. 한국어 속기시스템의 구성

한국의 폐쇄자막을 생성하는 속기시스템의 역사는 짧지만, 오래전부터 서비스를 실시한 미국, 일본 등에 필적하는 성능을 보이고 있다. 그림 5는 한국어 속기시스템의 구성을 나타낸 그림이다. 속기시스템은 네 명의 속기사들이 두 명씩 두 팀을 이루어 구성된다. 이 속기사들은 국가 공인 1급 자격증을 소지한 사람들로, 6개월에서 1년간의 폐쇄자막 입력을 위한 훈련을 거친 전문 속기사들이다. 보다 정확한 입력을 위해, 속기사들이

집중력을 유지할 수 있도록, 각 팀은 3초 간격으로 번갈아 가며 입력과 교정을 수행한다. 팀에서의 한명은 입력을 한명은 입력된 텍스트에 대한 교정을 위한 편집을 담당한다. 완료된 텍스트는 서버를 통해 방송국으로 송출된다. 한국어 속기시스템은 그림 5에서처럼 3대의 컴퓨터와 4대의 속기 단말기를 필요로 한다. 3대의 컴퓨터 중, 하나는 방송국으로의 송출을 담당하고, 나머지 2대는 폐쇄자막의 입력과 교정을 위해 사용된다. 현재 한국의 속기시스템은 생방송에 폐쇄자막을 삽입하는 실시간 시스템의 경우, 99.5%~99.9%의 정확도와 평균 지연 시간 2~3초 정도로 나타나 있다<sup>[4]</sup>. 속기시스템으로부터 생성된 폐쇄자막은 전문 속기사들에 의해 3초의 작업시간동안 처리된 일정한 양을 서버 컴퓨터를 통해 방송국으로 송출하게 되고, 시청자는 2~3초 후에 화면에서 폐쇄자막을 볼 수 있다.

2. 폐쇄자막의 시간할당

앞 절에서 살펴본 바와 같이, 실시간 속기시스템에서 생성된 일정한 양의 폐쇄자막은 비디오의 음성 발생 시점에서 2~3초의 지연시간을 갖는다. 즉, 실시간 한국어 속기시스템의 경우 속기사들이 생방송 중인 비디오의 음성을 듣고, 입력하고 교정한 후, 송출하는 시간은 거의 일정하게 유지된다고 볼 수 있다. 따라서, 폐쇄자막이 자막디코더에 수신되는 시간을 기록하여 폐쇄자막에 시간을 할당한 후, 일정한 지연시간만큼을 할당된 시간에서 다시 빼주면 비디오의 음성신호와 동기된 폐쇄자막을 얻을 수 있다. 물론 일반적인 녹화방송의 경우 사전에 폐쇄자막이 제작되어 삽입되어 있다면 그런 지연시간에 대한 추가적인 처리 없이 바로 자막디코더에 수신된 시간으로 동기를 맞출 수 있다. 이 실시간으로 처리되는 시간할당 방법은 다른 음성인식 등의 방법이 추가로 필요하지 않을 뿐더러, 빠르고 정확하게 시간 할당된 폐쇄자막을 얻을 수 있다는 장점을 갖는다. 그림 6은 폐쇄자막 디코더를 통해 수신되는 시간을 기록하기 위해 제작된 프로그램의 사용자 인터페이스를 보여주고 있다. 시간할당 프로그램은 그림 3에서와 같이 자막디코더와 RS-232C를 통한 직렬통신으로 데이터를 수신하고 있다.

폐쇄자막의 시간할당은 한국어의 어절단위로 수행된다. 지연시간에 대한 추가적인 처리를 수행할 때의 고려사항은, 폐쇄자막의 시간할당이 비디오 내의 음성 발생 시간보다 늦게 처리되지 않아야 한다는 것이다. 그

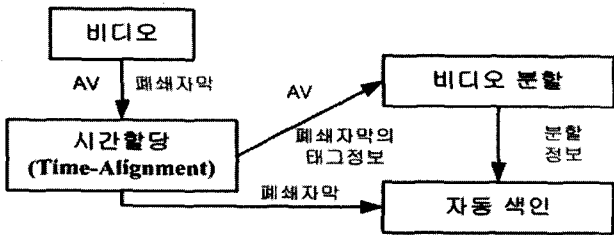


그림 8. 비디오 입력에서부터 자동색인 시스템까지의 구성  
Fig. 8. Schematic diagram from video to the automatic indexing system

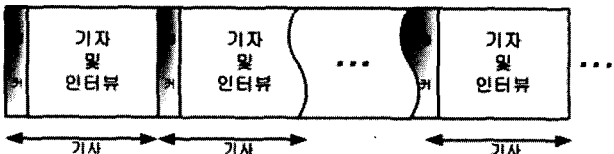


그림 9. 뉴스 비디오의 구조  
Fig. 9. Structure of news video.

렇게 되면 할당된 어절의 시간보다 음성이 앞에 동기되어, 그 시간 지점부터 비디오를 재생시 음성을 놓쳐버릴 수 있기 때문이다. 따라서 최대 지연시간인 3초를 전부 반영하여 지연시간에 대한 추가적인 처리를 수행하는 것이 보다 적절하다. 그림 7은 어절단위로 시간할당을 수행한 후, 데이터베이스에 저장된 폐쇄자막 정보를 보여주고 있다. 수신 시간의 단위는 1/100초로 기록한다. 폐쇄자막이 비디오의 오디오 및 영상 데이터와 동기되면, 폐쇄자막을 이용하여 효과적으로 비디오 검색을 수행할 수 있다.

#### IV. 자동 색인

폐쇄자막의 시간할당이 완료되면, 폐쇄자막의 문자열 매칭을 통하여 그 비디오에 대한 기본검색이 바로 가능하다. 그러나, 비디오를 분할하여 각 분할된 영역별로 검색을 수행하고 가장 높은 우선순위의 분할 영역을 제시하기 위해서는 키워드를 추출하기 위한 폐쇄자막에 대한 언어처리 및 가중치 부여가 필요하다.

본 논문에서는 실시간 한국에 속기시스템에서 생성된 폐쇄자막이 포함되어 있는 뉴스 비디오를 실험에 사용하였다. 뉴스 비디오의 경우 논리적인 비디오 분할이 가능하고, 폐쇄자막으로 표현되는 음성 데이터가 뉴스의 영상을 대표할 수 있을 정도로 정보량이 크기 때문에 폐쇄자막을 이용한 정보검색 시스템의 성능을 보이기에 가장 적합하다. 그림 8은 비디오 입력에서부터 자동색인과정까지의 구성도를 나타내고 있다.

MBC뉴스 윤능호입니다.

앵커: 올해부터는 닭, 오리 한 마리도 아무 데서나 잡을 수 없습니다. 도축장에서만 잡아야 하는데 아예 도축장이 없는 곳도 많아서 축산농가들만 줄지에 범법자가 되게 생겼습니다. 이진호 기자입니다.

기자: 충북 음성에서 오리농장을 하는 김만섭 씨는 오리를 도축하기 위해서 일주일에 두세 번씩 전라도 나주까지 가야 합니다. 올해부터는 오리도 지정된 곳에서만 도축을 해야 하기 때문입니다.

인터뷰: 가장 큰 문제점이 몰류비입니다. 도축비용도 비용이지만 몰류비가 도축비 대비 더 차지하고 있습니다.

기자: 그나마 정부에서 지정한 도축장에도 도축을 할 수 없습니다.

MBC뉴스 이진호입니다.

앵커: 노무현 대통령당선자는 오늘 전경련 토론회에서

기사 단위

그림 10. 폐쇄자막을 이용한 기사분할

Fig. 10. Segmentation of news video using the closed-caption.

##### 1. 폐쇄자막을 이용한 비디오 분할

뉴스 비디오에서는 정형적인 기사가 앵커 부분, 기자 및 인터뷰 부분으로 하나의 단위를 이루는 것이 일반적이다. 그림 9는 일반적인 뉴스 비디오의 구조를 보여주고 있다. 기존의 뉴스 검색시스템<sup>[5]</sup>에서는 그림 9와 같은 뉴스의 정형화된 구조적 특성을 이용하기 위하여 먼저 하나의 기사가 시작되는 앵커가 포함된 앵커 프레임을 추출하고 이를 경계로 기사단위로 뉴스 영상을 분할하고 있다. 그러나 앵커프레임과 유사한 프레임이 존재하거나, 처리의 오류로 인해 앵커프레임이 잘못 추출될 경우에는 분할 성능이 저하되고, 비디오 내의 모든 프레임을 처리해야 하기 때문에 많은 처리 시간을 필요로 하는 단점을 보이게 된다.

폐쇄자막은 비디오의 동영상 및 오디오 정보에 포함되어 있지 않은 화자의 구분을 위한 태그 정보가 삽입되어 있다. 한국의 뉴스 비디오의 경우 '앵커:', '기자:', '인터뷰:'의 태그들이 존재한다. 따라서 기존의 앵커 프레임을 기준으로 기사별로 분할하는 대신 앵커 태그와 각 태그의 구조, 항상 뉴스명과 기자의 이름을 호출하는 기사의 끝맺음 형식 등을 기준으로 분할 작업을 수행하면 앵커 프레임을 잘못 추출하여 발생하는 분할의 오류를 해결할 수 있다. 그림 10은 폐쇄자막의 구조로부터 기사를 분할하는 예를 보여주고 있다. 본 논문에서는 그림 10에서 처럼 시작되는 '앵커:' 태그와 기사를

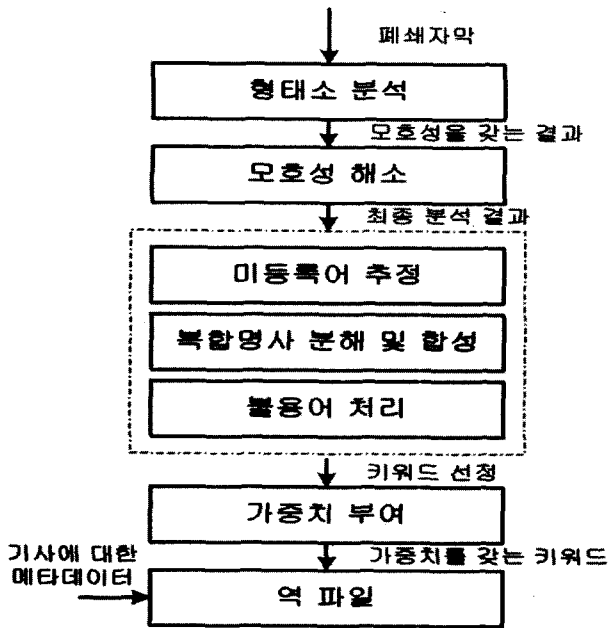


그림 11. 키워드 추출 시스템의 구성도  
 Fig. 11. Schematic diagram of the keyword extraction system.

마무리하는 기사 멘트 다음의 ‘앵커.’ 태그 사이를 기사 단위로 분할한다.

2. 키워드 추출 시스템

기사가 분할되면 각 기사의 폐쇄자막에서 키워드를 추출한다. 기사 내의 폐쇄자막은 명사를 대상으로 키워드를 추출하기 위해, 어절별로 자체 개발한 한국어 형태소 분석기<sup>[6]</sup>로 입력된다. 그림 11은 키워드 추출 시스템의 구성을 나타낸다.

형태소 분석 결과는 여러 개의 분석 후보를 갖게 되며, 이러한 후보들에서 명사로 분류되는 어휘를 색인어 후보로 전부 추출하게 되면 불필요한 어휘가 상당수 포함되게 된다. 따라서 결과의 모호성을 해소하기 위해 [6]에서의 인접한 어절의 형태소 및 태그를 이용하여 설정된 모호성 해소 규칙을 적용한다. 이렇게 해서 얻어진 명사에 불용어(stopword) 목록을 적용하는 것만으로도 어느 정도의 성능을 보이지만, 보다 정확한 키워드 추출을 위해 한국어의 자동 색인에서 필요한 처리를 더하였다. 모호성이 해소된 그 결과로부터 사전에 등록되어 있지 않은 미등록어를 추정하고, 붙여 쓴 복합 명사의 분해 및 띄어 쓴 복합명사의 결합에 관한 처리를 수행하기 위해 한글 복합 명사 패턴 정보를 사용하였다. 그런 다음 불용어를 제거하여 검색에 사용할 키워드를 최종적으로 선정하였다<sup>[6]</sup>.

선정된 각 키워드는 역문헌 빈도의 변형식인 식 (1)에 따라 가중치를 부여한다.

$$W_{ik} = (f_{ik}/AF) \times \{ \log_2(n) - \log_2(AF) + 1 \} \quad (1)$$

$W_{ik}$  는  $i$  번째 기사에서 키워드  $k$  가 갖는 가중치를 의미하며,  $f_{ik}$ 는  $i$  번째 기사에서 키워드  $k$  가 출현한 빈도수를 의미한다.  $AF$ 는 색인 대상이 되는 뉴스 비디오 내에서 특정한 단어  $k$ 가 출현한 기사의 빈도수를 나타내고  $n$ 은 현재 색인 작업 중인 뉴스 비디오 내의 전체 기사의 수를 의미한다. 여기서 데이터베이스 내의 전체 기사의 수 대신 가중치 산정에서 현재의 색인 대상이 되는 뉴스 비디오 내의 전체 기사의 수를 기준으로 하는 이유는 날마다 새로 제작되는 뉴스기사가 데이터베이스에 계속 추가 삽입되는 매우 유동적인 특징을 갖기 때문이다. 식 1은 전체 기사의 크기나 분석 대상이 되는 기사의 길이, 키워드의 출현 빈도를 고려하여 단순 빈도인 출현 빈도 값을 표준화시킨 상대빈도에 해당하며, 특히 식별력이 높은 키워드에 가중치를 크게 부여하도록 [7]과 [8]의 역문헌 빈도 변형식에 기사빈도를 다시 나눠주었다. 이와 같은 가중치 부여를 통해 사용자의 질의에 대한 결과에 우선 순위를 부여할 수 있다. 키워드가 선정되고 가중치가 산정되면 각 키워드는 해당 기사의 방영날짜, 기사의 구분번호, 가중치, 뉴스명, 방송국명, 앵커명, 기자명 등의 정보를 갖는 역파일을 구성한다. 이러한 메타데이터는 색인 관리자가 직접 입력하도록 구성하였다.

3. 내용기반 검색을 위한 대표프레임에서의 칼라 정보 추출

기사 단위로 분할한 후, 기사를 대표할 수 있는 프레임을 선정한다. 본 논문에서는 하나의 기사 내에서 ‘앵커’, ‘기자’, ‘인터뷰’ 태그가 각각 처음 등장하는 시간과 동기된 영상 프레임을 대표 프레임으로 선정하였다. 이러한 대표프레임에 대해 내용기반 검색이 가능하도록 MPEG-7의 색상 구조 기술자를 이용하여 특징 정보를 추출하여 색인 데이터베이스에 저장한다. 사용자는 키워드 외에도 예제 기반 질의(query by example)를 통해 비디오 내의 이미지 프레임들을 검색할 수 있으며 비디오 검색과 연동시킬 수 있다.

CSD(color-structure descriptor)<sup>[9]</sup>는 색상 히스토그램 정보와 색상의 구조정보를 동시에 표현하는 색상 특징

기술자이다. 색상 히스토그램 정보는 특정한 색을 가진 이미지 픽셀의 상대 빈도수로 표현되고, 색상 구조 정보는 특정한 색을 가진 이미지 샘플을 포함하는 구조적 엘리먼트(SE, structuring element)의 상대 빈도수로 표현된다. 이러한 색상 구조 정보는 색상 히스토그램 정보가 표현하지 못하는 색상의 공간적 분포를 표현할 수 있다. CSD의 특징정보는 전체 이미지 영역에 대하여 SE를 이동시키면서 SE내에 포함되어 있는 이미지 샘플의 색상에 해당되는 히스토그램 빈을 증가시키는 과정을 통하여 얻을 수 있다. 일반적으로 SE는 8×8 블록을 사용한다. 예를 들어 그림 12에서 히스토그램 빈이 8개라고 가정하고, 8×8 SE 내부에는 C1, C3, C7의 히스토그램 빈에 해당되는 색상이 포함되어 있다. 이때 SE 내에 포함되어 있는 색상에 해당되는 히스토그램 빈 C1, C3, C7을 각각 하나씩 증가시킨다. 이러한 과정이 SE가 전체 영역을 이동하면서 이루어지게 되면 지역적으로 집중되어 있는 색상은 해당 히스토그램 빈을 다수에 걸쳐 증가시키게 된다. 결과적으로 CSD는 색상 히스토그램 정보를 포함하는 동시에 색상의 공간적인 밀집도를 나타내는 구조 정보가 포함된 색상 특징정보를 표현하게 된다. 본 논문에서는 각 프레임마다 256차원의 특징정보를 추출하여 내용기반 검색에 사용하였다.

V. 비디오 검색

방대한 양의 비디오 데이터에 대한 재사용은 매우 중요하다. 따라서 데이터에 대한 효과적인 관리와 함께

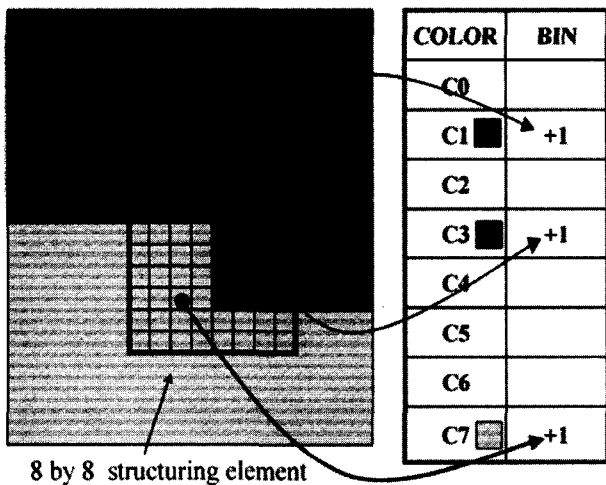


그림 12. 색상 구조 기술자의 추출 과정  
Fig. 12. The extraction procedure of the color-structure descriptor.

사용자가 원하는 데이터를 보다 빠르고 정확하게 찾기 위한 검색 시스템은 필수적이다<sup>[10]</sup>. 본 논문에서는 검색을 기본검색과 고급검색으로 나누어 설계한다. 기본검색은 사용자가 입력한 키워드 질의와 문자열 매칭되는 위치를 기준으로 검색한다. 고급검색은 폐쇄자막으로부터 키워드를 추출하고 가중치 색인 과정을 거쳐 분할된 기사 단위의 검색이 가능하며, 각 분할된 기사의 대표 프레임들을 대상으로 내용기반 검색을 수행할 수 있는 다양한 기능을 포함한다.

1. 기본검색

기본검색은 사용자가 질의로 입력한 키워드에 해당하는 비디오 내의 동영상 및 오디오 구간을 검색할 수 있다. 질의어가 입력되면 그림 7에서 나타낸 데이터베이스를 이용하여 비디오 내에서 음성과 동기된 시간을 먼저 얻는다. 이 시간정보에 비디오 인코딩시 설정한 초당 프레임수를 식 2처럼 곱하면 재생해야할 구간의 시작 프레임 번호를 얻을 수 있다.

$$F_{start} = (t - r) \times m \tag{2}$$

여기서, 시작 프레임 번호는  $F_{start}$ , 음성과 동기된 시간이  $t$ 이며  $m$ 은 인코딩할 때 설정된 초당 프레임 수이다.  $r$ 은 지연시간으로 폐쇄자막에 대한 시간할당 작업에서 지연시간에 대한 부가적인 처리를 하지 않았을 경우, 실시간 속기시스템이 적용된 생방송의 경우 3을, 녹화방송의 경우 지연시간이 없도록 미리 동기가 맞추어져 있으므로 0으로 설정하여 사용한다. 기본검색에서도 뉴스 기사의 정형화된 구조적 특성을 이용하여 질의어가 포함되어 있는 기사에 대한 기사 단위의 검색이 가능하다. 먼저 질의어에 해당하는 키워드를 기준으로 이전에 수신된 가장 가까운 ‘앵커’ 태그를 찾는다. 그런 다음 그 동기된 시간을 찾아 식 2를 적용하여 재생하면 쉽게 기사 단위의 검색을 수행할 수 있다. 또한 부가적인 기능으로 사용자가 직접 재생 시작 시간을 입력하여, 원하는 시간 위치부터 재생 가능하도록 하는 기능도 포함되어 있다. 그림 13은 기본검색에 대한 사용자 인터페이스를 보여주고 있다. 이러한 기본검색은 그림 6의 시간할당 프로그램에 의해 폐쇄자막의 어절 단위로 비디오의 동영상 및 오디오 정보와의 동기된 시간만 할당되면 바로 서비스를 이용할 수 있는 장점이 있다.

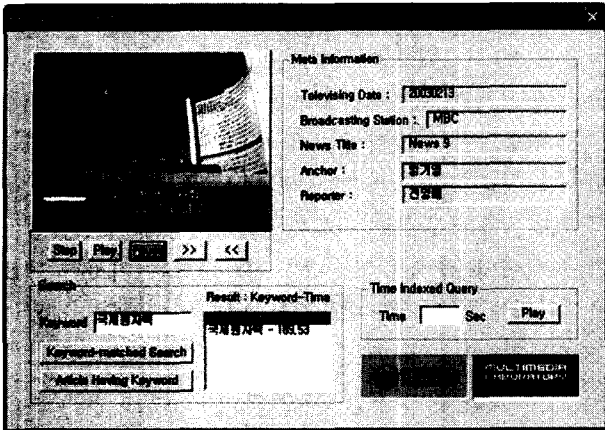


그림 13. 기본 검색을 위한 사용자 인터페이스  
Fig. 13. User interface of basic search.

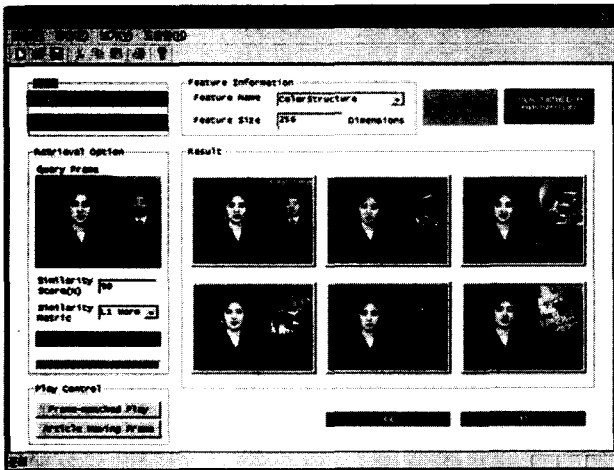


그림 14. 칼라 구조 기술자를 이용한 대표프레임에 대한 내용기반 검색  
Fig. 14. Content-based retrieval using the color-structure descriptor.

2. 고급검색

본 논문에서 적용 대상으로 하고 있는 뉴스 비디오의 경우, 하루에도 수많은 기사가 새로 만들어진다. 이러한 기사들을 효과적으로 검색하기 위해서는 식별력이 높은 키워드의 추출과 중요한 키워드에 대한 가중치 부여가 필요하며, 각 기사와 연관된 기자명, 방송날짜 등의 메타 정보의 사용이 요구된다. 고급검색에서는 사용자가 입력한 질의에 대해 각 키워드에 대한 가중치 계산을 통해 전체 데이터베이스를 검색한 결과들에 우선순위를 부여하여 제공한다. 사용자는 물론 검색한 기사 내에서 키워드와 부합되는 동영상 및 오디오 구간을 검색하기 위해 기본검색을 다시 적용할 수도 있다. 자동색인 과정에서 얻어진 역파일을 사용하여 선호하는 방송국이나 뉴스 종류별로 검색할 수도 있으며, 날짜별 검색이 가

능하다. 또한 사용자가 방송국의 데이터 저장 관리자라면 앵커나 기자와 같은 정보도 찾고자 하는 기사를 찾아주는데 효과적으로 이용될 수 있을 것이다. 그림 14는 MPEG-7의 색상 구조 기술자를 이용하여 대표 프레임들을 예제기반 질의를 통해 검색하는 인터페이스를 보여주고 있다. 여기서 찾고자 하는 대표프레임을 선택하면 사용자는 대표프레임에 해당하는 기사에 관련된 메타데이터 정보를 확인할 수 있으며, 대표프레임을 시작프레임으로 하여 기사를 재생하거나, 해당 기사 전체를 재생할 수 있다.

폐쇄자막 정보를 사용하는 이러한 검색 방법은 키워드 자체가 단순한 문자열 형태 이상의 의미를 갖는 정보이기 때문에 그림 13의 결과와 같이 동기된 동영상 및 오디오 구간도 그 키워드의 의미와 연관성이 높아 사용자의 개념적인 요구를 수용할 수 있는 장점이 있다. 또한 메타데이터 정보와 함께 기존의 내용기반 검색 방법도 하나의 질의 수단으로 제공함으로써 검색 효율을 극대화하였다. 특히 뉴스 비디오에 적용할 경우 폐쇄자막 데이터는 음성 정보를 나타내고 있기 때문에 정보량이 매우 커 효율성 또한 더욱 큼을 확인할 수 있다.

VI. 성능 평가

검색 시스템의 적합성에 대한 평가는 기준이 되는 검색 결과의 객관적인 답을 얻기 어렵고, 그 적합한 정도도 사용자의 주관적인 판단에 의해 크게 좌우되므로 임의로 설정한 답과의 단순한 비교는 크게 의미가 없다. 본 논문에서 제안하는 비디오의 음성데이터에 등장하는 키워드를 기반으로 검색하는 기본검색과 이를 확장하여, 키워드를 통한 가중치 검색 및 기사의 메타데이터 검색, 내용기반 검색을 지원하는 고급검색은 정확한 폐쇄자막의 시간할당 및 검색단위로 구분하기 위한 정확한 비디오 분할이 성능을 좌우한다. 따라서, 시간할당에 대한 정확도 평가와 분할 결과의 정확도 평가를 통해 시스템의 성능을 평가하고자 한다. 본 논문에서는 2003년 2월에 방송된 MBC 9시 뉴스 28일 분량을 실험에 사용하였다. 9시 뉴스는 하루의 기사를 종합하는 대표 뉴스 프로그램으로 대표성이 높고, 폐쇄자막이 포함되어 있어 성능을 평가하기에 적합하다.

폐쇄자막과 비디오의 동영상 및 오디오 데이터와의 동기를 맞추는 시간할당에 대한 성능평가는 폐쇄자막으



표 1. 폐쇄자막의 시간할당의 정확도

Table 1. Precision of time-alignment of closed-caption.

성능평가용 전체 키워드 추출 개수	1000개
시간할당 성공 키워드의 개수 (3초 이내)	997개
	99.7%

표 2. 뉴스 분할의 정확도

Table 2. Precision of news segmentation

2003년 2월 전체 기사 수 (스포츠, 날씨 제외)	756개
분할 성공 기사 수	735개

로 등장하는 키워드를 임의로 1000개 선정하고, 각 키워드에 동기된 오디오 구간에 해당 키워드가 존재하는지 여부를 평가하였다. 존재 여부를 평가하는 오디오 구간은 동기시점부터 3초 이내까지이며, 그 결과는 표1과 같다.

시간할당이 실패한 키워드 3개는 동기시점에 해당 키워드에 대한 음성이 재생을 시작한 직후로, 음성만을 청취하여서는 정확한 키워드 인식이 불가능하였다. 이 결과는 3개의 키워드가 3초의 최대지연시간보다 더 큰 지연시간을 갖고 있기 때문으로 분석되었다. 그러나 99.7%에 해당하는 대부분의 키워드에 대한 지연시간은 3초 이내에 들어 성공적인 시간할당이 수행되었음을 확인할 수 있었으며, 또한 실패시에도 오디오 구간과 매우 근접한 동기 결과를 보이고 있어, 기사의 전체내용을 전달하는 데에는 큰 문제가 없기 때문에 기본검색시 충분한 성능을 보장할 수 있다.

IV장에서 설명한 바와 같이, 본 논문에서 적용하고 있는 뉴스비디오는 대부분 정형적인 구조를 갖고 있다. 특히, 하루의 뉴스를 대표하고 있는 9시 뉴스의 경우에는 더욱더 그러하다고 볼 수 있다. 분할 성능에 대한 평가를 위해 2003년 2월에 방영된 뉴스에 폐쇄자막의 태그를 이용하는 제안하는 분할 방법을 적용하여 MBC 홈페이지(<http://news.imbc.com/vas/desk.html>)에 구축되어 있는 결과와 비교하였다. 그 결과는 표 2와 같다.

앵커가 기사를 간단히 기사를 소개하고 기자이름을 호출한 후, 기자가 기사를 전달하고 기자가 속한 방송사와 기자의 이름을 호출하며 하나의 기사를 끝맺는 구조는 국내의 모든 방송사에서 공통적으로 적용되고 있는 일반적인 구조이다. 실험에서는 뉴스에서 스포츠와 날씨를 제외한 뉴스를 대상으로 하여 뉴스의 정형적인 구조를 갖는 기사를 먼저 분할하고, 정형적인 구조를

갖지 않는 뉴스 뒷부분의 기사를 단신 종합기사로 묶어 분할하였다. 분할에 실패한 기사들을 분석한 결과, 중대한 사건이 발생하여 뉴스의 정형적인 구조를 벗어나는 경우로 예를 들면, 스튜디오 안에서 앵커와 전문가가 특정 주제로 대화를 나누는 부분이 이에 해당한다. 그러나 그와 같은 경우도 대부분 앞의 기사나 뒤의 기사에 포함되어 연속된 기사로도 판단할 수 있으므로 크게 문제되지는 않는다. 분할 성능을 좀 더 높이기 위해서는 폐쇄자막의 태그들 외에도 앵커 프레임 등의 다양한 부가적인 정보를 함께 사용하면 효과적일 수 있다.

이상의 결과로, 사용자가 입력하는 키워드 질의를 통해 원하는 동영상 및 오디오 구간을 재생할 수 있으며, 또한 해당 기사를 찾아 정확하게 기사 단위로 검색할 수 있음을 확인하였다. 사용자가 입력하는 키워드는 단순한 문자열이 아닌, 키워드 자체에 상당한 의미가 내포되어 있다고 볼 수 있으며, 따라서 이 키워드를 통한 비디오 검색 방식의 사용으로 메타데이터 검색, 내용기반 검색 등과 함께 복합적인 질의가 가능하기 때문에 사용자의 개념적인 검색 요구에 보다 부합하는 결과를 얻을 수 있다.

## VII. 결 론

본 논문에서는 지식기반 비디오 검색을 위한 한국어 폐쇄자막의 자동색인 방법과 색인된 데이터베이스를 이용한 다양한 검색방법을 제안하였다. 한국어 속기시스템이 생성하는 폐쇄자막의 정확도와 지연시간을 파악하여, 폐쇄자막의 동영상 및 오디오와 동기된 시간을 자막디코더의 수신시간만을 사용하여 성공적으로 할당하였다. 색인 과정에서는 폐쇄자막에 포함된 태그 정보를 사용하여 비디오를 효과적으로 분할하였고, 고급검색을 위해 폐쇄자막에 대한 형태소분석 및 분석결과의 모호성해소를 수행하고, 미등록어 처리, 복합명사 처리, 불용어 제거의 과정을 거쳐 색인어를 선정하였으며, 기사에 대한 메타데이터 정보를 같이 저장하였다. 제안한 방법은 기본검색에서 키워드를 통해 동기된 동영상 및 오디오의 구간을 찾아 쉽게 검색할 수 있으며, 키워드가 포함된 기사에 대한 재생도 가능하였다. 고급검색에서는 색인 데이터베이스를 이용하여 기사별 가중치 검색, 방송국명, 뉴스명, 날짜, 기차명, 앵커명에 따른 검색이 가능하고 대표프레임에 대한 내용기반 검색이 가능하였다.

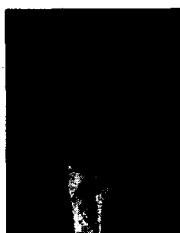
실험을 통해 음성을 그대로 표현하는 폐쇄자막 정보가 사용자의 개념적인 요구에 부합하는 비디오 검색 결과를 제공할 수 있음을 확인하였으나, 현재 한국에서는 청각장애인을 위한 폐쇄자막이 들어있는 방송이 10~20%에 머무르고 있는 실정이다. 청각 장애인을 위한 복지차원 뿐만 아니라, 방송 데이터의 효율적인 저장과 검색을 위해서도 분야에 상관없이 서비스의 비율을 높일 필요가 있다. 향후에는 적용 대상을 음성의 정보량이 역시 높은 다큐멘터리 비디오로 우선 확장할 예정이며, 이후 드라마나 영화에도 적용할 예정이다.

**참고 문헌**

[1] Vittorio C., Lawrence D.B, *Image Databases*, Jon Wiley & Sons, Inc., pp. 261 - 279, 2002.  
 [2] Howard D.W., Takeo K., Michael A.S., Scott M.S, "Intelligent Access to Digital Video: Informedia Project," *IEEE Computer (Digital Library Initiative special)*, Vol. 29, No.5, pp. 46 - 52, 1996.  
 [3] Alexander G.H., Michael J.W., "Informedia: News-on-Demand Multimedia Information Acquisition and Retrieval," *Intelligent Multimedia Information Retrieval*, AAAI Press/The MIT Press, pp. 215 - 240, 1997.

[4] Han-Woo K., "Stenographic System for Caption Service of T.V Broadcasting," *Proceedings of TAO International Workshop on TV Closed Captions for the Hearing Impaired People*. Nov. 20, pp. 31-35, 1999.  
 [5] Inderjeet M., David H., Mark T.M., Morgan G., "Towards Content-Based Browsing of Broadcast News Video," *Intelligent Multimedia Information Retrieval*, AAAI Press/The MIT Press, pp. 241 - 258, 1997.  
 [6] Jung-Won C., *An Automatic Indexing System for Intelligent Information Retrieval*, Master's thesis, Hanyang University, pp. 26-28, 1998.  
 [7] William B.F., Ricardo B.-Y., *Information Retrieval: Data Structure and Algorithms*, Prentice-Hall, 1992.  
 [8] Salton, G., *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*, Addison Wesley Publishing, 1989.  
 [9] MPEG-7 Visual Experimentation Model (XM), Ver. 10.0, *ISO/IEC/JTC1/SC29/WG11, Doc N4062*, 2001.  
 [10] Wei Q., Lie G., Hao J., Xiang-Rong C., Hong-Jiang Z., "Integrating visual, audio and text analysis for news video," *Proc. of Int. Conf. on Image Processing*, Vol. 3, pp. 520 - 523, 2000.

저 자 소 개



**조 정 원**(정회원)  
 1996년 인천대학교 정보통신 공학과 학사 졸업  
 1998년 한양대학교 전자통신 공학과 석사 졸업  
 2004년 한양대학교 전자통신 전파공학과 박사 졸업

<주관심분야: 멀티미디어 정보검색, 멀티미디어 콘텐츠 처리 및 보안, 자연언어처리, 웹기반 시스템, E-Learning>



**최 병 옥**(정회원)  
 1973년 한양대학교 전자공학과 학사 졸업  
 1978년 일본 경응의숙(KEIO) 대학 전기공학과 석사 졸업  
 1981년 일본 경응의숙(KEIO) 대학 전기공학과 박사 졸업

현 한양대학교 정보통신대학 정보통신학부 교수  
 <주관심분야: 영상처리, 멀티미디어 공학>



**정 승 도**(학생회원)  
 1999년 한양대학교 전자·전자 통신·전파공학과 학사 졸업  
 2001년 한양대학교 전자통신전파 공학과 석사 졸업  
 현재 한양대학교 전자통신전파 공학과 박사과정 재학 중

<주관심분야: 컴퓨터비전, 생체인식, VR/AR IBR>