

일반논문-09-14-4-10

디지털방송 자막데이터를 이용한 멀티미디어 응용 서비스 연구

김정연^{a)}, 남제호^{b)†}

A Study on Multimedia Application Service using DTV Closed Caption Data

Jung-Youn Kim^{a)} and Jeho Nam^{b)†}

요 약

본 논문에서는 DTV 자막방송 데이터의 응용을 통한 부가가치 서비스 활용 방안을 연구한다. DTV 자막방송(Closed-Captioning)이란 방송프로그램의 대사를 문자화된 자막으로 보여주는 서비스로서 장애인, 노약자, 외국인 등 정보소외계층의 방송접근권 확대를 통한 정보격차(digital divide) 해소를 목적으로 한다. 국내에서는 2007년 6월에 디지털TV 자막 방송 표준규격이 완료되었으며, 2008년 4월부터 ‘장애인 차별금지 및 권리구제 등에 관한 법률’에 의한 모든 방송서비스의 자막방송 의무화로 인하여 현재 대부분의 지상파방송에서는 자막방송 서비스를 제공하고 있다. 본 논문에서는 DTV 자막방송 데이터의 응용을 위하여 지상파방송의 MPEG-2 Transport Stream으로부터 자막데이터를 추출하는 방법을 설명하고, 추출된 자막과 시간정보를 이용하여 PC 환경의 멀티미디어 재생기에서 사용가능한 자막파일의 생성 방법을 제안한다. 그리고 생성된 자막파일을 이용한 방송콘텐츠의 시간적 구간분할 방법을 제안한다. 실험에서는 자막파일 생성 및 구간분할 S/W의 구현을 통하여 제안된 방법의 유용성을 확인하였으며, 다양한 자막데이터 응용 기능의 제시를 통하여 자막데이터의 응용 가능성을 검증하였다.

Abstract

In this paper, we study on making a use of value-added services using DTV closed caption data. Note that Closed-Captioning service helps to bridge “digital divide” through extending broadcasting accessibility of a neglected class such as hearing-impaired person and foreigner. In Korea, DTV Closed Captioning standard was developed in June 2007, and Closed Captioning service should be provided by an enforcing law in all broadcasting services in April 2008. Here, we describe the method of extracting a caption data from MPEG-2 Transport Stream of ATSC-based digital TV signal and generating a caption file using the extracted caption data and time information. In addition, we present the segmentation method of broadcasting content using caption file. Experimental results show that implemented S/W tool provides the feasibility of the proposed methods and the usability of closed caption for a variety of data application service.

Keyword : DTVC, DTV 자막, 자막파일, 구간분할, ATSC, Segmentation

a) (주)휴맥스 부설연구소 개발본부

R&D Department, HUMAX Co., Ltd.

b) 한국전자통신연구원(ETRI) 방송미디어연구부

Broadcasting & Telecommunications Media Research Department, ETRI

† 교신저자 : 남제호(namjeho@etri.re.kr)

※ 본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 문화콘텐츠산업 기술지원사업의 일환으로 수행하였음. [2-09-1205-001-10987-11-001, 국가표준콘텐츠식별체계를 이용한 방송콘텐츠 유통 서비스 프레임워크 기술 개발]

지상파 디지털 TV 자막방송 국내의 표준기술에 대해 도움을 주신 KBS 이동준, MBC 이상규, SBS 김성환, EBS 송주호 연구원 외 방송 4사 기술연구소 관계자 여러분께 감사의 뜻을 전합니다.

· 접수일(2008년12월4일),수정일(1차:2009년5월20일,2차:6월29일),게재확정일(2009년7월21일)

I. 서론

디지털방송의 보편화에 힘입어 일반 사용자의 방송콘텐츠 접근과 소유가 용이해지고 있다. MPEG-2 TS(Transport Stream)^[1]로 전송되는 디지털 방송스트림에는 오디오·비디오 신호 외에 PSI(Program Specific Information^[1]), PSIP(Program and System Information Protocol^[2]) 등의 다양한 데이터가 함께 다중화되어 있다. 그리고 또 다른 방송데이터로서 DTV 자막 서비스를 위해 제공되는 자막데이터가 있다. DTV 자막방송(Closed-Captioning)이란 방송프로그램의 대사를 문자화된 자막으로 보여주는 서비스로서 장애인, 노약자, 외국인 등 정보소외계층의 방송접근권 확대를 통한 정보격차(digital divide) 해소를 목적으로 한다. 국내에서는 2007년 6월에 디지털TV 자막 방송 표준규격이 완료되었으며^[3], 2008년 4월부터 ‘장애인 차별금지 및 권리구제 등에 관한 법률’에 의해 모든 방송서비스의 자막방송 의무화로 인하여 현재 대부분의 국내 지상파방송에서는 자막 방송 서비스를 제공하고 있다. 자막데이터는 디지털방송의 전송규격인 MPEG-2 TS에 다중화되어 있으며^[4], 수신기에서 자막을 재생하기 위해서는 별도의 자막 추출 및 재생 기능이 필요하다. 한편, 이러한 PSI, PSIP, 자막 등의 디지털 방송데이터는 다양한 형태의 부가가치 서비스에 활용될 수 있으며, 특히 DTV 자막데이터는 본래의 기능 외에 방송콘텐츠의 내용 검색 및 색인 등에 이용될 수 있다.

본 논문에서는 DTV 자막데이터의 응용을 통한 부가가치 서비스 활용 방안을 제안한다. 우선 MPEG-2 TS로부터 자막데이터를 추출하고, 추출된 자막과 시간정보를 이용하

여 PC 환경의 멀티미디어 재생기에서 사용가능한 자막파일의 생성 방법을 제안한다. 그리고 생성된 자막파일을 이용한 방송콘텐츠의 시간적 구간분할(temporal segmentation) 방법을 제안한다. 구간분할은 방송콘텐츠의 장르에 따른 자막데이터의 분석을 통하여 수행되며, 본 논문에서 구간분할을 수행하는 대상 방송콘텐츠의 장르는 뉴스, 시사토론, 그리고 드라마이다. 실험에서는 자막파일 생성 및 구간분할 기능을 갖춘 S/W의 구현을 통하여 제안된 방법의 유용성을 확인하며, 다양한 자막데이터 응용 기능의 제시를 통하여 자막데이터의 응용 가능성을 검증한다.

본 논문의 구성은 다음과 같다. II장에서는 DTV 방송콘텐츠로부터 자막데이터의 추출 및 자막파일의 생성 방법을 설명한다. 그리고 III장에서 자막파일을 이용한 방송콘텐츠의 구간분할 방법을 설명하며, IV장에서 제안된 자막파일 생성 및 구간분할 기능과 자막데이터 응용 기술의 구현을 통하여 제안된 방법의 유용성을 확인한다. 끝으로 V장에서 결론을 맺음으로써 본 논문을 마친다.

II. 자막데이터 추출 및 자막파일 생성 방법

본 장에서는 자막데이터를 추출하고, 추출된 자막을 규격화된 자막파일로 변환하는 방법을 설명한다. 자막파일 규격은 국내외에서 자막파일 규격으로 가장 널리 이용되고 있는 SAMI (Synchronized Accessible Media Interchange) 표준을 사용한다^[5]. 그림 1은 제안된 DTV 자막 추출 및 SAMI 자막파일 생성기의 블록도를 나타낸다.

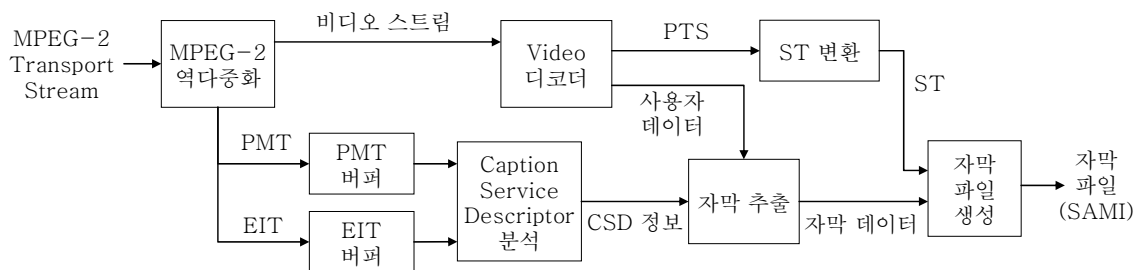


그림 1. 자막데이터 추출 및 자막파일 생성 블록도
 Fig. 1. Block diagram of extracting closed caption data and creating caption file

표 1. Caption Service Descriptor의 비트 스트림 구문^[3]
Table 1. Syntax of the Caption Service Descriptor^[3]

Syntax	No. of Bits	Format
caption_service_descriptor() { ... number_of_services	5	uimsbf
for (i=0; i<number_of_services; i++) { language	8*3	uimsbf
... korean_code	1	bslbf
... } }		

1. 디지털방송 자막의 추출

본 절에서는 DTV 자막데이터의 추출 과정을 설명한다. 자막 추출 대상은 지상파 DTV 방송스트림의 전송 단위인 MPEG-2 TS이며, MPEG-2 Systems^[1], PSIP 표준규격^[2] 및 국내외 DTV 자막방송 표준규격^[3,6], 그리고 ATSC (Advanced Television Systems Committee) A/53 표준규격^[4]을 참고하여 본 연구의 자막 추출과 해석과정을 수행한다.

1.1 Caption Service Descriptor 해석

자막의 추출에 앞서서 자막 서비스 서술자(Caption Service Descriptor; CSD)에 대한 해석이 필요하다. CSD는 PSI의 PMT(Program Map Table) 또는 PSIP의 EIT(Event Information Table)에 존재하는 서술자로서 자막의 유형과 속성을 기술하며, 표 1은 CSD의 비트 스트림 구문을 나타낸다. 표 1의 language는 자막의 언어를 나타내는 3 byte 코드이다. 각 언어의 코드는 ISO 639.2/B에 정의되어 있으며^[7], 한국어의 경우 'kor'로 표현된다. korean_code는 국내의 자막방송 규격에만 정의되어있는 필드로서, 자막 언어가 한글인 경우 완성형(0)인지 혹은 유니코드(1)인지를 나타낸다. 그 밖의 모든 필드에 대한 분석이 완료되면, 이후에 전송되는 자막은 CSD의 정보에 따라서 해석된다.

표 2. user_data 영역의 구문^[4]
Table 2. Syntax of the user_data^[4]

Syntax	No. of Bits	Format
user_data() { user_data_start_code	32	bslbf
ATSC_identifier	32	bslbf
user_data_type_code	8	uimsbf
if (user_data_type_code == '0x03')		
cc_data()		
... next_start_code()		
}		

1.2 비디오 스트림 추출

디지털 방송스트림의 표준규격인 MPEG-2 Systems에 정의된 TS(Transport Stream)의 패킷 식별자(Packet Identifier; PID)를 통하여 해당 TS의 페이로드(payload)가 어떤 데이터(예, 비디오, 오디오 등)인지를 알 수 있다. 비디오 스트림의 추출은 PAT(Program Association Table)와 PMT의 해석에 의해 비디오 스트림의 PID를 얻어냄으로써 수행된다. 우선 188 byte 패킷 단위의 MPEG-2 TS에서 PID가 0x0000인 PAT를 찾아낸 후, program_map_PID 필드의 분석을 통해 PMT의 PID를 알아낸다. PMT의 PID를 이용해 PMT를 찾아낸 후 stream_type 필드가 0x02인 elementary_PID 필드의 분석을 통해서 비디오 스트림의 PID를 찾아낼 수 있다.

1.3 자막데이터 추출

추출된 비디오 스트림은 PES(Packetized Elementary Stream)으로 구성되어 있으며, 비디오 PES 내의 사용자 데이터(Picture user data)의 구조는 표 2와 같다. 사용자 데이터에는 자막데이터를 넣도록 규정된 자막데이터(cc_data) 필드가 정의되어 있으며 그 구조는 표 3과 같다. 자막데이터 필드 중 cc_data_1과 cc_data_2는 자막데이터의 첫 번째 byte와 두 번째 byte를 나타내며, cc_count의 개수만큼의 자막데이터를 구성할 수 있다.

위의 과정을 통하여 구성된 자막데이터는 패킷 계층에 해당한다. 이후에 뒤따르는 서비스 계층, 코딩 계층, 그리고 해석 계층의 분석을 통하여 최종적인 자막데이터와 자막의

표 3. 자막데이터의 구분³⁾
Table 3. Syntax of Captioning Data³⁾

Syntax	No. of Bits	Format
cc_data() { ... for(i=0 ; i<cc_count ; i++) { ... cc_data_1 cc_data_2 } ... }	8 8	bslbf bslbf

구성에 대한 정보를 얻을 수 있다^{3,6)}.

그림 2(a)는 자막이 포함된 뉴스 실험영상, 그림 2(b)는 실험영상으로부터 추출된 자막의 일부분을 나타낸다.

2. 자막파일 생성

앞선 절의 과정을 통하여 추출된 자막데이터는 규격화된 자막파일로 변환된다. 본 논문에서 사용하는 자막파일 규격인 SAMI 파일은 HTML 기반의 자막파일이다⁵⁾. SAMI 파일 생성을 위해서는 재생되는 영상과의 동기화 시간 (Sync Time; ST)과 각 ST에 재생되는 자막의 적절한 배치가 필요하다. 아래의 과정을 통하여 결정된 ST와 연결된 자막은 자막파일(*.smi) 생성을 위하여 SAMI 파일 규격이 적용된다.

2.1 동기화 시간 계산

SAMI 파일구조에는 기본적으로 자막이 재생되는 milli-second(ms) 단위의 동기화 시간 정보가 포함된다. DTV 방송 자막데이터는 비디오 스트림에 포함되어 있기 때문에 비디오 스트림 PES의 헤더에 포함되어 있는 PTS(Presentation Time Stamp)를 SAMI 파일의 자막 재생 시간 정보로 활용 가능하다. PTS는 PES 헤더에 위치한 33 bit의 필드로서 PES의 재생시간을 나타낸다. 단위는 시스템 클럭 주파수 단위이며, 본 논문에서는 SAMI 파일의 동기화 시간 단위로 환산하기 위한 방법을 식 (1)과 같이 제안한다.

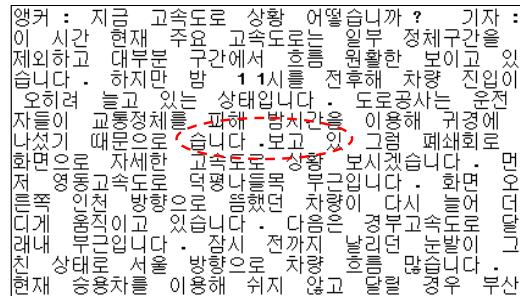
$$ST = (PTS/90) - (PTS_{start}/90). \quad (1)$$

PTS를 초 단위로 환산하기 위해서는 90kHz로 나누는 연산이 필요하다¹⁾. 그러나 90kHz로 나누면 소수점 이상의 값이 초 단위로 얻어지기 때문에 소수점 이상의 값을 SAMI의 ms 단위로 얻기 위하여 PTS를 90Hz로 나눈다. PTS_{start} 는 첫 번째 PES 헤더의 PTS를 의미한다.

그림 2(b)의 점선으로 표시된 부분에서 보여지듯이, PES의 전송 순서대로 추출하면 자막의 순서가 다르게 추출될 수 있다. 따라서 PES는 비디오 프레임의 디코딩(decoding) 순서대로 전송 및 저장되기 때문에 자막 추출 시에는 그림 3과 같이 PTS의 순서, 즉 프레임 재생시간 순서대로 정렬하여 자막을 추출해야한다.



(a)



(b)

그림 2. 자막 추출 결과: (a) 자막이 포함된 뉴스 영상, (b) 추출된 자막의 일부분

Fig. 2. The result of extracting closed caption: (a) News program containing closed caption and (b) A part of extracted closed caption

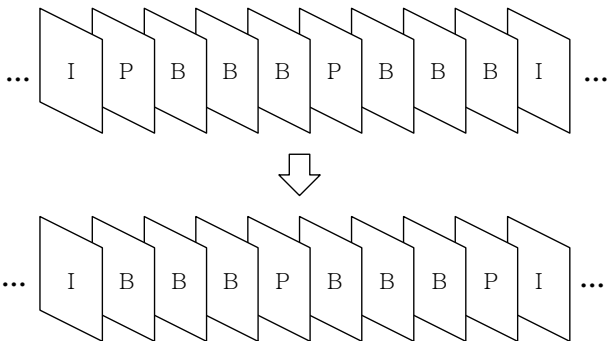


그림 3. 비디오 프레임의 PTS 순서에 따른 정렬.
Fig. 3. Arrangement of video frames in order of PTS

2.2 자막 연결 및 배치

추출된 자막을 완전한 단어나 문장의 형태로 배열하기 위해서는 상황에 따라서 다수의 PES에서 추출된 자막을 문장 단위 혹은 일정 길이로 연결하는 작업이 필요하다. TV 화면에 나타나는 자막의 행이나 열의 수를 결정하기 위한 기준의 하나로서, DTV 자막방송표준의 해석계층에 정의되어 있는 Command Descriptions 중 하나인 DefineWindow를 이용할 수 있다⁶⁾.

DefineWindow의 row count와 column count(이하 row/column count)는 각각 화면에 나타내는 행과 열의 수를 나타내며, row lock과 column lock(이하 row/column lock)은 row/column count에서 명시된 값을 화면 출력 시에 고정된 값으로 사용하는지 여부를 나타낸다. 즉 row/column lock의 값이 Yes(1)로 설정되었을 때에는 명시된 row/column count에 맞춰서 화면에 자막이 재생되어야 하지만, No(0)로 설정되었을 때에는 row/column count의 값이 화면 재생 시에 절대적이지 않다는 의미이다.

제안된 방법에서는 자막의 유연한 배치를 위하여 row/column lock이 No(0)로 설정된 경우만을 고려하며, 이때 row/column count는 각 ST에 자막이 배치되는 최대 길이의 기준으로 사용한다. 그림 4는 각 ST에 대응되는 자막의 연결 과정 순서도이다.

특수문자는 자막에 따라서 CSD의 korean_code 필드에 명시된 한글자막의 종류(완성형 또는 유니코드)와 다르게 1 byte의 ASCII 코드일 수도 있으므로 시스템 설계 시에 이를 반영해야 한다.

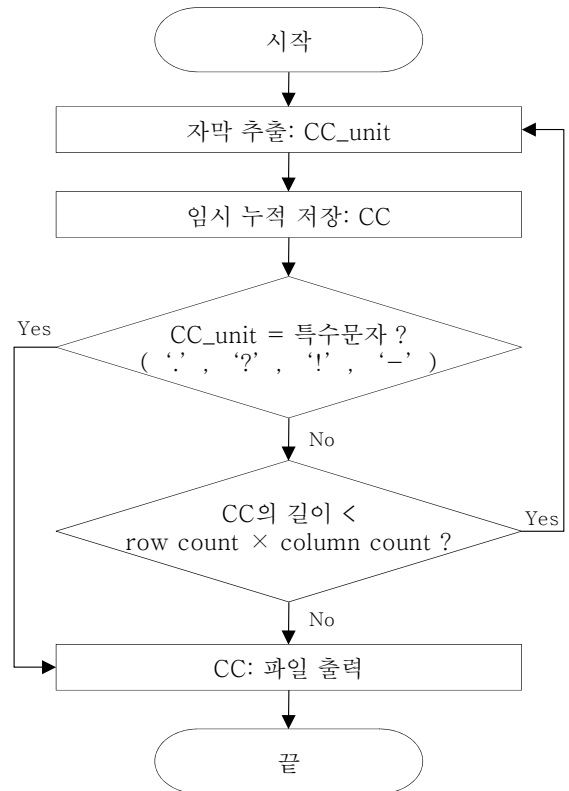


그림 4. 자막 연결 순서도
Fig. 4. Flow-chart of linking closed-caption

자막의 연결 과정을 거치면서 서로 분리되어 있던 자막을 하나로 합침에 따라서 각각 자막데이터에 대응하는 다수의 ST 중에서 연결된 자막을 대표하는 하나의 ST를 결정해야 한다. 제안된 방법에서는 연결된 자막에서 중간(median) 자막의 ST를 최종 ST로 결정한다. 그림 5(a)는 각 PES에서 추출, 변환된 ST와 자막데이터를 나타내며, 그림 5(b)는 위의 과정을 거쳐서 최종 선택된 ST와 연결된 자막의 예를 나타낸다.

III. 자막 기반 방송콘텐츠 구간분할

앞선 과정에서 생성된 자막파일은 멀티미디어 재생기에서 자막을 보여주는 기본적인 기능 외에 비디오 검색 및 색인 등 다양한 응용 데이터로서 활용될 수 있다. 본 절에서

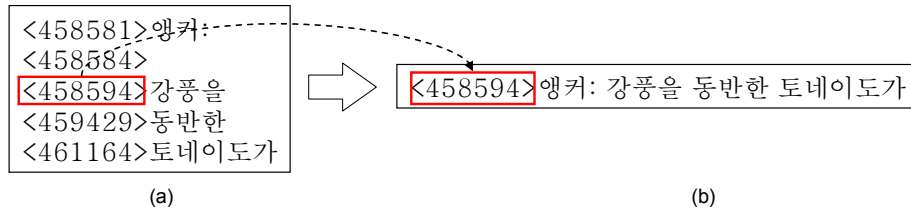


그림 5. 최종 ST의 결정 및 연결된 자막: (a) 추출된 ST와 이에 대응하는 자막데이터, (b) 선택된 최종 ST와 연결된 자막
 Fig. 5. Decision of final ST and linked closed caption: (a) Extracted ST and closed caption data corresponding to each ST and (b) Selected final ST and linked closed caption

는 자막데이터를 이용한 방송콘텐츠의 시간적 구간분할 (temporal segmentation) 방법을 설명한다.

본 절에서 시간적 구간분할을 수행하는 대상 방송콘텐츠의 장르는 뉴스, 시사토론, 그리고 드라마이다. 기존연구에서는 국내 자막방송의 특성을 기술하고, 자막데이터를 이용한 뉴스의 검색 및 구간분할 방법을 제안하였다⁸⁾. 그러나 구간분할의 대상 장르가 뉴스에 한정되어 있기 때문에 제안된 방법을 다른 장르의 구간분할에 적용하는 것은 불가능하다. 국내 방송의 경우, 자막데이터는 장르마다 상이한 특징정보를 가지고 있기 때문에 방송콘텐츠의 장르에 따라서 서로 다른 구간분할 방법을 적용해야 한다. 본 절에서는 각 방송콘텐츠 장르에 따른 자막데이터의 특성과 함께 비디오 구간분할 방법에 대해서 설명한다. 시간적 구간분할을 통하여 구간의 시작시간, 재생시간, 그리고 해당 구간의 자막데이터를 얻을 수 있다. 제안된 구간분할 방법은 미리 추출된 자막데이터를 이용해서 이루어지기 때문에 기존의 비디오 프레임 기반의 장면분할 방법과 비교해서 분할 처리속도가 매우 빠르다.

1. 뉴스

일반적으로 뉴스의 구간분할 단위는 하나의 기사이다. 국내 방송 뉴스의 자막은 실제 대사에 없는 ‘앵커:’, ‘기자:’, 그리고 ‘인터뷰:’ 등 화자를 구분할 수 있는 일종의 태그(tag)를 포함하며, 뉴스는 일반적으로 기사를 마무리하는 일정한 대사가 존재한다. 그림 6(a)는 뉴스 자막데이터의 예이며, 뉴스 기사는 다음과 같은 기준을 통해서 구분될 수 있다.

㉠ ‘앵커:’가 나오고 다음에 ‘앵커:’가 나오면 이는 하나의 독립된 뉴스 기사이다.

㉡ ‘앵커:’가 나오고 다음에 ‘앵커:’가 나오기 전에 ‘기자:’가 나오면 기자의 이름을 저장하고, 이후에 “[방송사 이름]뉴스 [기자이름]입니다.” 라는 문장이 나오면 여기까지는 하나의 뉴스 기사로 구분한다.

㉠은 하나의 뉴스 기사 내에 앵커의 대사만 포함된 형태에, ㉡은 앵커가 뉴스 기사의 개요 설명을 하고나서 기자의 세부적인 보도 내용이 따르는 경우 또는 앵커와 기자가 해당 뉴스 기사에 관하여 대화식으로 설명을 주고받는 뉴스 형식에 적용된다. 이와 같은 뉴스 자막데이터의 특징정보 분석을 통하여 방송사 이름과 기자 이름을 비교적 쉽게 얻을 수 있다.

2. 시사토론

국내 방송의 경우 화자전환(speaker change) 시에 자막데이터에 하이픈('-') 기호를 삽입하여 청각장애인이나 외국인 이 화자전환을 인식할 수 있도록 돕는다. 그림 6(b)는 하이픈이 포함된 시사토론 자막데이터의 예이다. 시사토론 프로그램의 경우, 토론자 각자의 의견을 일정 시간동안 제시하기 때문에 다른 방송 장르에 비해서 비교적 화자전환의 시간 간격이 길며, 화자의 전환에 기반한 구간분할이 효과적이다. 따라서 제안된 방법은 화자전환 표시인 하이픈과 최소구간간격의 설정을 통하여 시사토론 방송콘텐츠의 시간적 구간분할을 수행한다. 최소구간간격이란 일종의 구간분할 기준으로, 최소구간간격이 정해지면 그 안에 발생

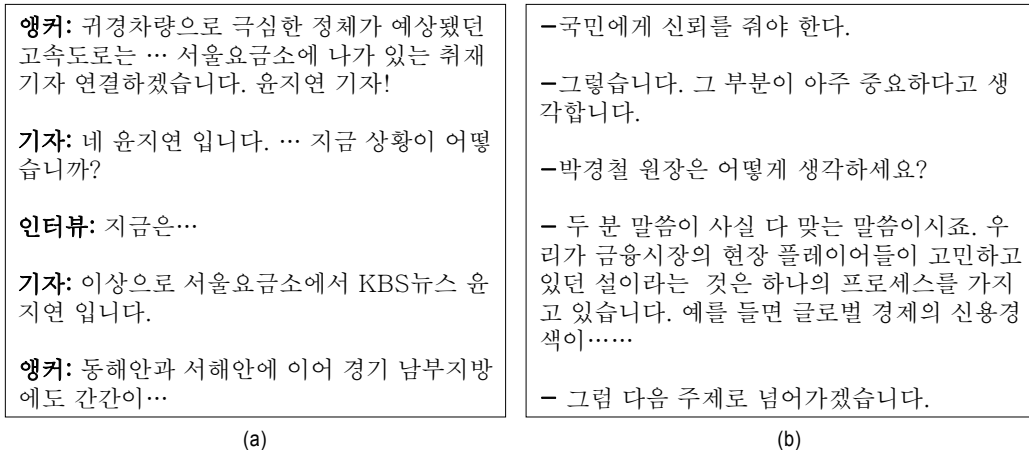


그림 6. 자막데이터의 예: (a) 뉴스 자막데이터, (b) 시사토론 자막데이터
 Fig. 6. Examples of closed caption data: (a) News and (b) Debate

하는 화자전환은 하나의 구간으로 인식한다. 예를 들어, 최소구간간격을 20초로 설정하였을 경우, 20초 내로 발생하는 화자전환은 무시하고 하나의 연속된 구간으로 인식하며, 해당 구간의 시작시간 기준으로 20초가 지난 이후 나타나는 화자전환 표시부터는 새로운 구간으로 인식한다. 최소구간간격의 설정은 사용자의 선호에 따라서 임의로 설정 가능한 변수로서, 사용자가 원하는 최소한의 구간길이를 설정하는 기능으로 활용될 수 있다.

3. 드라마

드라마의 경우에도 하이픈 기호를 이용한 시간적 구간분할이 가능하다. 그러나 앞서서 제시한 시사토론에서의 장면분할 방법은 화자전환이 빈번한 드라마의 특성상 효율적이지 못하다. 따라서 본 논문에서는 자막 기반의 드라마 구간분할 방법을 다음과 같이 제안한다.

우선 하이픈 기호로 시작되는 화자전환된 자막데이터를 받았을 경우 다음 식 (2)의 계산을 통하여 화자전환된 자막의 예상되는 동기화시간(Expected_ST)을 얻는다.

$$\text{Expected_ST} = (NW \times 60) / (\alpha + \beta) + \text{PreST}. \quad (2)$$

NW은 바로 이전 ST에 해당하는 자막의 단어 개수이며,

α 와 β 는 각각 1분당 말하는 단어 수와 화자전환 대기시간을 의미한다. α 와 β 는 사용자 선호에 따라서 임의로 설정이 가능한 변수이다. α 가 클수록 1분당 말하는 단어 수를 크게 적용하기 때문에 계산되는 Expected_ST의 값이 작아진다. β 는 α 로 인하여 얻어지는 시간에 더하여 다음 자막이 발생하기까지 대기하는 시간을 설정하는 변수이다. 이 두 값과 얻어지는 자막의 단어 수를 통하여 해당 자막의 재생 시간(duration)과 다음 자막이 발생하기까지 대기하는 시간의 합을 예상한다. 여기에 바로 이전 자막의 ST을 의미하는 PreST를 더함으로써 현재 화자전환된 자막의 ST를 예측하는 것이다. 계산된 Expected_ST와 현재 화자전환된 자막의 ST를 비교하여 ST가 Expected_ST보다 클 경우에 현재의 화자전환된 자막을 새로운 구간으로 인식한다. α 와 β 의 조절은 분할된 구간의 개수에 영향을 미치는 변수로서, α 가 크거나 β 가 작을수록 더 많은 구간으로 분할될 수 있다. 그림 7은 $\alpha=80$, $\beta=6$ 으로 설정했을 때의 드라마 구간분할의 예를 나타낸다. 그림 7의 계산 1에서는 ST (289756ms)가 Expected_ST(296321ms)보다 작기 때문에 앞선 자막과 같은 구간으로 인식하며, 계산 2에서는 화자전환된 자막데이터의 ST가 더 크기 때문에 새로운 구간의 시작으로 판단한다.



그림 7. 자막데이터를 이용한 드라마 구간분할의 예($\alpha=80, \beta=6$)

Fig. 7. An example of drama segmentation using closed caption data ($\alpha=80, \beta=6$)

IV. 성능평가 실험

본 장에서는 자막 추출 및 자막파일 생성 S/W와 구간분할 및 다양한 자막데이터 응용 기능을 갖춘 S/W의 구현을 통하여 제안된 방법의 유용성을 검증한다. 각각의 구현은 Microsoft Visual C++ MFC를 이용하였으며, 실험대상은 HDTV 수신카드를 통하여 녹화한 지상파 DTV 방송콘텐츠이다. HDTV 수신카드는 디비코 Fusion HDTV7 RT Gold 모델을 사용하였으며, 각각의 입력은 앞서 언급했듯이 디지털 방송스트림인 MPEG-2 TS이다.

1. 자막 추출 및 자막파일 생성기 구현

앞서 설명한 과정에 따라서 DTV 방송 자막데이터의 추

출 및 SAMI 자막파일 생성 S/W를 구현하였다. 그림 8은 구현된 S/W의 실행화면이다. 그림 8의 우측 하단에 위치한 ‘자막 싱크 설정’을 통하여 녹화방송과 생방송을 구분하여 자막을 추출할 수 있다. 뉴스와 같은 생방송에서 실시간 자막을 제공할 경우 국내 자막방송은 약 2~4초의 지연(delay)이 발생하므로^[9] 이를 고려하여 지연을 통한 동기화 시간의 보정이 필요하다. 생성된 SAMI 파일은 그림 8의 ‘생성된 SAMI 파일’에서 보는 바와 같이 규격에 맞춰서 SAMI 파일이 생성되었음을 알 수 있다. 그림 9는 생성된 SAMI 자막파일을 적용하여 일반 PC 환경의 범용 멀티미디어 재생기에서 실행한 실험영상의 일부분이다. 사용한 멀티미디어 재생기는 곰 플레이어이며^[10], 자막이 영상의 동기에 맞춰서 재생됨을 확인할 수 있다.

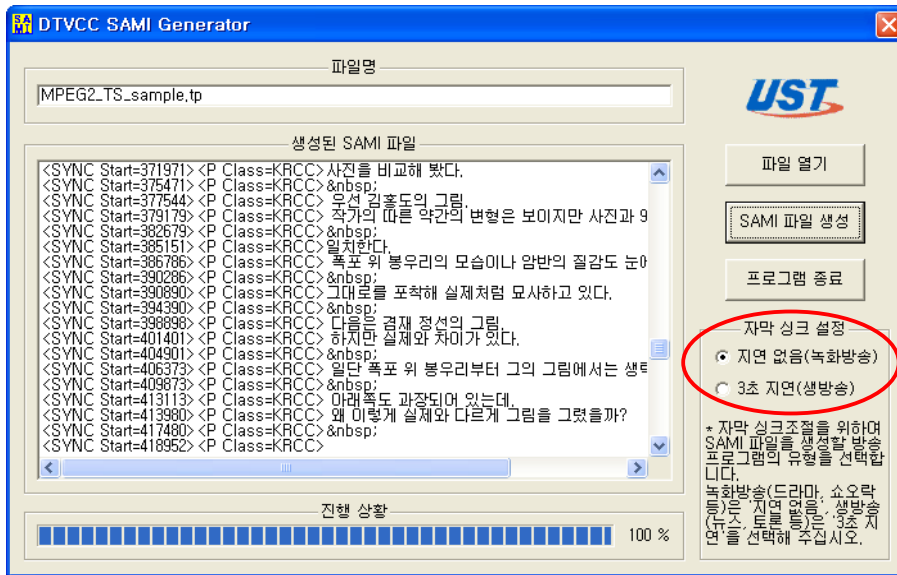


그림 8. 자막 추출 및 SAMI 파일 생성기의 GUI
 Fig. 8. GUI of extracting closed caption and SAMI file generator



그림 9. 생성된 자막파일을 일반 멀티미디어 재생기에 적용한 실험영상
 Fig. 9. Test video applying generated caption file to general multimedia player

2. 자막데이터 기반 구간분할 S/W 구현

제안된 구간분할 기법과 자막데이터의 유용성을 확인하기

위하여 구간분할 기능 및 자막데이터 응용 기능을 갖춘 S/W를 구현하였다. 그림 10은 구현된 S/W의 실행화면이다. 구현된 재생기는 구간분할 기능 외에도 자막데이터를



그림 10. 구간분할 S/W의 GUI
Fig. 10. GUI of temporal segmentation S/W

이용한 자막 보여주기, 키워드 기반 방송콘텐츠 장면 검색 기능 등을 제공하며, 다이렉트쇼(DirectShow) 기술^[11]을 이용하여 PC용 멀티미디어 재생기가 갖춘 일반적인 동영상 재생 기능을 제공한다.

2.1 시간적 구간분할 기능

시간적 구간분할 기능을 수행하기 위하여 구현된 S/W에서는 우선 방송콘텐츠의 장르를 지정할 수 있다. 본 연구에서 구간분할 대상으로 정한 장르는 뉴스, 시사토론, 그리고 드라마이며, 각각의 장르에 따른 구간분할에 사용되는 파라미터를 정할 수 있도록 사용자 인터페이스를 구현하였다. 시사토론의 경우에는 최소구간간격, 드라마의 경우에는 분당 말하는 단어 수(α)와 화자전환 대기시간(β)을 정할 수 있다. 그림 11은 드라마 구간분할의 예를 나타낸다. 실험에서는 드라마의 α 와 β 를 각각 80개와 6초로 정하였으며, 하단에 구간분할의 결과로서 해당 구간의 키프레임 썸네일(thumbnail), 시작시간, 재생시간, 그리고 자막데이터를 제공한다. 또한 구현된 S/W는 다이렉트쇼 기술을 이용한 랜덤액세스(random access) 기능을 제공하여 해당 구간의 시작위치로 비디오의 이동이 가능하며, 사용자의 주관적 판단에 따른 파라미터의 조절을 통하여 인터랙티브(interac-

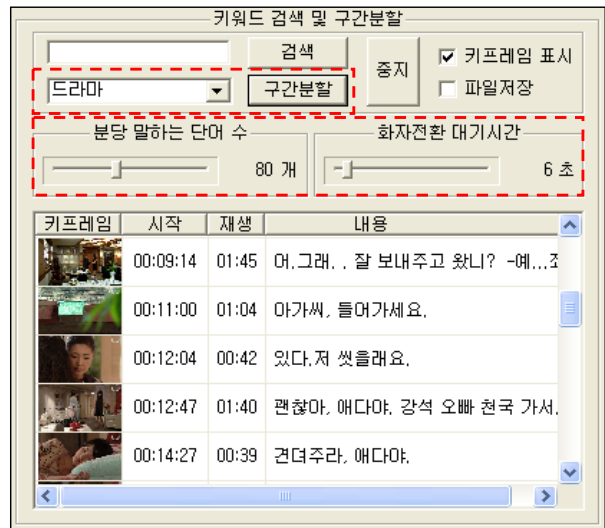


그림 11. 드라마 구간분할의 예
Fig. 11. An example of segmentation of drama/soap

tive)한 방식으로 분할된 구간의 개수를 설정할 수 있다. 본 절에서 수행하는 구간분할의 성능평가는 엄밀한 객관적 측정이 불가능하기 때문에 아래와 같이 개념적으로 의미를 갖는 시간적 구간을 검증 목적의 비교대상으로 정한 후, 구간분할 결과의 적절성을 평가한다.

- 뉴스: 기사 단위
- 시사토론 : 개별 토론자의 발언 및 토론 주제의 변화
- 드라마 : 장면의 변화(자막 정보를 기준으로 일정 시간 이상의 대사가 존재하지 않으면 서로 다른 장면으로 구분함)

실험결과를 통하여, 제안된 자막 기반의 구간분할 과정이 매우 빠르게(on-the-fly) 수행되었으며, 상기의 주관적 비교기준에 따라 구간분할이 적절하게 이루어졌음을 확인할 수 있었다. 또한 적절한 구간분할을 수행하고자 파라미터 설정을 달리하여 반복수행 할 때, 시간적 지연 없이 빠른 반복처리가 가능함을 확인하였다.

구현된 S/W에서 구간분할의 결과는 그림 12와 같이 파일 형태로 기록, 제공될 수 있다. 그림 12의 <StartTime>은 ‘시:분:초’ 단위로 표시한 구간의 시작 시간 정보이며, <Caption>은 해당 구간의 자막데이터를 나타낸다. 이러한 텍스트 기반의 파일은 MPEG-7^[12], TV-Anytime^[13]과 같은 표준규격 기반의 메타데이터 형태로 변환이 용이하며, 이러한 메타데이터를 지원하는 기기에서 유용하게 활용될 수 있다.

```

<StartTime>00:00:00
<Caption>
-전화가 정말로 고장이 났어요. 네, 새로 사라고 할게요. 이사장님 안 계시는데. 오늘도 아무도 안 계세요. 어보세요? 어보세요! 강 회장 아드님이 화 많이 냈잖나 봐요, 이사장님.. (노크소리).
-사부인.
-안 여사님!
-어떻게..
...
<StartTime>00:01:05
<Caption>
-네.
-월 그렇게 싸웠어. 입맛도 없는데.
-그래도 잡수셔야죠. 애다는 병원에 있죠?
-네, NS그룹이요?
...
<StartTime>00:04:47
<Caption>
-제가 어떻게 하면 되겠습니까?
-조용히 떠나줘. 애다 곁에서. 그게 네가 할 수 있는 전부야. 얼마면 돼? 필요한 액수 얘기해 봐.
    
```

그림 12. 구간분할 정보 파일의 예
 Fig. 12. An example of segmentation information file

2.2 자막 재생 기능

그림 13은 구현된 S/W의 자막 재생 기능을 나타내고 있다. 실험결과를 통하여 영상의 음성과 동기화를 이루며 자막이 출력됨을 확인하였다. 이 기능은 그림 9의 일반 멀티



그림 13. 자막 재생 기능
 Fig. 13. A function of displaying closed caption

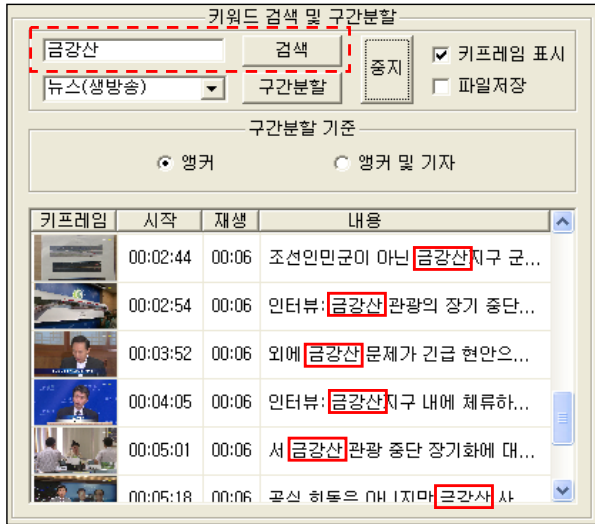


그림 14. 키워드 기반 장면 검색 기능의 예
 Fig. 14. An example of keyword-based scene search function

미디어 재생기에서 자막의 재생과 달리 별도의 자막보기 창을 통해 이전에 재생된 자막까지 모두 볼 수 있는 기능을 제공한다. 이러한 기능은 현재 자막과 이전 영상의 지나간 자막을 동시에 제공함으로써 보다 효과적인 정보의 전달이 가능하다.

2.3 키워드 기반 방송콘텐츠 장면 검색 기능

키워드 기반의 장면 검색 기능 역시 자막데이터를 검색함으로써 수행될 수 있다. 그림 14는 장면 검색 기능의 사용 예를 나타내고 있다. 좌측 상단에 키워드를 입력하고 ‘검색’ 버튼을 누르면, 구현된 S/W에서는 입력된 키워드를 자막데이터에서 검색하고, 키워드가 발견되면 해당 자막의 키프레임 썸네일(thumbnail)과 시작 시간정보를 제공한다. 또한 구간분할에서와 같은 랜덤액세스 기능을 제공하여 키워드가 발견된 위치로 비디오의 이동이 가능하다. 이 검색 기능은 방송콘텐츠에서 화자가 말하는 내용을 검색할 때 유용하며, 특히 상대적으로 정보성이 높은 뉴스나 시사토론 프로그램의 콘텐츠 브라우징(browsing) 및 탐색에 유용하게 사용될 수 있다.

V. 결 론

본 논문에서는 디지털방송 자막데이터의 응용 서비스 방안을 연구하였다. 우선 DTV 자막방송의 자막데이터 추출방법과 추출된 자막데이터를 이용하여 자막파일을 생성하는 방법을 설명하였으며, 자막파일의 자막데이터와 시간 정보를 이용하여 방송콘텐츠의 시간적 구간분할 방법을 제안하였다. 실험에서는 자막 추출 및 자막파일 생성기의 구현을 통하여 생성된 자막파일이 일반 PC 멀티미디어 재생기에서 영상과 동기화하여 자막 재생이 가능함을 확인하였으며, 구간분할 S/W의 구현을 통하여 제안된 구간분할 방법의 유용성을 검증하였다. 또한 자막 재생, 키워드 기반 비디오 장면 검색, 구간분할 정보 파일 저장 기능의 구현을 통하여 DTV 자막방송 데이터의 다양한 응용 가능성을 확인하였다. 생성된 SAMI 표준을 따르는 자막파일은 별도의 기기가 필요없이 일반 PC의 멀티미디어 재생기에서 재생이 가능하기 때문에 청각장애인의 방송접근권 확대에 기여할 수 있을 것이며, 다른 언어로의 번역을 통하여 외국인에게 제공될 수 있을 것으로 기대된다.

제안된 디지털방송 자막데이터를 이용한 멀티미디어 응용 서비스는 자막 정보를 이용하기 때문에 검색 등에서 비교적 정확하고 풍부한 정보를 제공한다. 또한 완전히 텍스트 기반이므로 고속처리가 가능하며, 이는 구간분할 시 더욱 유용하게 이용될 수 있다. 예를 들어 사용자는 적절한 구간분할을 수행하고자 파라미터 설정을 달리하여 반복수행 할 때, 시간적 지연 없이 빠른 반복처리가 가능하다. 또한 자막파일, 구간분할 정보, 장면 검색 결과 등의 정보는 HTML 및 XML과 같은 다양한 정보문서 형태로의 변환이 용이하다. 특히 시간적 구간분할 정보는 MPEG-7^[12] 또는 TV-Anytime^[13] 표준규격 메타데이터로 변환하여 이를 지원하는 기기에서 활용될 것으로 기대된다. 향후 제안된 방법을 응용하여 DTV 자막데이터를 통한 다양한 부가가치 정보 생성 및 메타데이터 서비스에 대해서 연구하고자 한다.

참 고 문 헌

- [1] "Information technology - Generic coding of moving pictures and associated audio information: Systems," ISO/IEC International Standard 13818-1, Dec. 2000.
- [2] "Program and system information protocol for terrestrial broadcast and cable (Revision C) with amendment No. 1," Doc. A/65C, Advanced Television Systems Committee, Jan. 2006.
- [3] "지상파 디지털 TV 자막방송," TTA, TTAS. KO-07.0050, Jun. 2007.
- [4] "Digital television standard, Part 2 - MPEG-2 video system characteristics," Doc. A/53, Part 4:2007, Advanced Television Systems Committee, Jan. 2007.
- [5] "Understanding SAMI 1.0," Microsoft Developer Network (MSDN), <http://msdn2.microsoft.com/en-us/library/ms971327.aspx>, Feb. 2003.
- [6] "Digital television (DTV) closed captioning," EIA-708-B, Dec. 1999.
- [7] "Code for the representation of names of languages - Part 2: alpha-3 code," ISO 639.2, as maintained by the ISO 639/Joint Advisory Committee (ISO 639/JAC), <http://www.loc.gov/standards/iso639-2/iso639jac.html>, Dec. 2008.
- [8] J. Cho, S. Jeong, and B. Choi, "News video retrieval using automatic indexing of Korean closed-caption," LNCS 3683, pp. 694 - 703, Aug. 2005.
- [9] 한국CAS속기협회, <http://www.casi.co.kr/>, Dec. 2008.
- [10] 고프레이어, <http://gomtv.gretech.co.kr/>, Dec. 2008.
- [11] "DirectShow," Microsoft Developer Network (MSDN) http://msdn.microsoft.com/en-us/library/ms78332_VS.85.aspx, Nov. 2008.
- [12] "Multimedia content description interface-Part 5: multimedia content description schemes," ISO/IEC International Standard 15938-5, May 2002.
- [13] The TV-Anytime Forum, <http://www.tv-anytime.org>, Dec. 2008.

저 자 소 개



김 정 연

- 2007년 2월 : 광운대학교 컴퓨터공학과 학사
- 2009년 2월 : 과학기술연합대학원대학교(UST) 이동통신 및 디지털방송공학과 석사
- 2009년 2월 ~ 현재 : (주)휴맥스 부설연구소 연구원
- 주관심분야 : 워터마킹, 디지털방송기술, 파일포맷, 비디오 암호화, DTV SW개발, 셋탑박스 시스템



남 제 호

- 1992년 2월 : 홍익대학교 전기제어공학과 학사
- 1996년 12월 : University of Minnesota, Electrical Engineering 석사
- 2000년 12월 : University of Minnesota, Electrical Engineering 박사
- 2001년 2월 ~ 현재 : 한국전자통신연구원(ETRI) 방통융합미디어연구부 선임연구원
- 2007년 1월 ~ 현재 : 과학기술연합대학원대학교(UST) 이동통신 및 디지털방송공학과 부교수
- 주관심분야 : 신호처리, 디지털방송기술, 멀티미디어 보호관리, MPEG