화면해설방송 콘텐츠 저작 기술

장인선, 임우택, 안충현
한국전자통신연구원 방송통신미디어연구부문
{jinsn, wtlim, hyun}@etri.re.kr

Descriptive Video Service Contents Authoring Technique

Inseon Jang, Wootaek Lim, ChungHyun Ahn
Electronics and Telecommunications Research Institute

요 약

본 논문에서는 시각장애인의 TV 시청 보조기술인 화면해설방송(Descriptive Video Service)에 있어 다양한 편의 기술을 적용하여 저작자로 하여금 편리하고 경제적으로 화면해설방송물을 제작할 수 있도록 하는 저작 기술을 제안한다. 제안하는 방법은 화면해설 대본 작성의 편의를 위한 화면해설 삽입 추천 기술과 성우 음성 더빙을 대체할 수 있는 TTS 기반의 합성음성 기술, 또한 마스터 오디오와 합성음성 화면해설을 믹싱하기 위한 오디오 믹싱 기술을 포함한다. 마지막으로, 제안하는 기술의 구현 예를 제시한다.

1. 서론

화면해설방송이란 시각장애인들이 TV 프로그램 및 영화와 같은 미디어에 접근할 수 있도록 하는 서비스로써 화면을 볼 수 없는 시각장애인들을 위해 상황 변화적 요소와 자막, 그래픽 등의 시각적 요소들을 음성으로 설명하여 프로그램 내용을 이해하도록 도와준다. 2011 년 방송통신위원회에서 제정한 “장애인방송 편성 및 제공 등 장애인방송 접근권 보장에 관한 고시”에 따라 장애인에게 안정적이고 체계적인 방송접근권을 보장하기 위하여 중앙지상파 방송사와 보도 및 종합편성 채널에서는 각각 2014 년과 2016 년까지 전체 방송 프로그램의 10%를 화면해설방송으로 편성하도록 의무화 되었으며 케이블 및 IPTV 사업자에도 2016 년까지 5~7%의 의무편성이 규정되었다[1]. 그 결과, 국내 화면해설방송의 편성 비율은 지속적으로 증가하는 추세이나 화면해설 제작과정의 특성상 시간과 비용이 상당히 소요되기 때문에 다큐멘터리 등 특정 장르에 치우쳐 제작되는 현실이다.

화면해설 콘텐츠의 저작 편의성을 높이기 위해 2010 년대 초부터 다양한 저작 기술/도구가 개발되어 왔다. 이 저작 도구들은 화면해설 텍스트 입력/편집 등 단순한 화면해설 대본 저작기능을 제공하거나 텍스트 기반 포맷으로의 출력만을 지원하는 등 화면해설 콘텐츠 저작 전반에 대해 편의성 제공에는 한계가 있었다[2][3][4].

본 논문에서는 효율적인 화면해설 저작 방법을 제안한다. 제안하는 방법은 MPEG-2 TS (Transport Stream) 등 다양한 동영상 파일로부터 비디오 분석을 통해 화면전환 구간을 감지하고 오디오와 자막 정보를 분석하여 비 대사 구간을 검출하여 저작자에게 화면해설 삽입 구간을 추천함으로 효율적인 화면해설 대본 작성을 가능하게 한다. 또한 저작자가 입력한 화면해설 텍스트를 TTS (Text-to-Speech)를 통해 음성으로 변환하고 변환된 화면해설 합성음성을 마스터 오디오와 믹싱하여 출력함으로써 화면해설 콘텐츠의 1 인 저작을 가능하게 한다.

본 논문의 구성은 다음과 같다. 2 절에서는 제안하는 화면해설 저작기술의 특징을 설명하고 3 절에서는 제안한 방법을 포함하는 화면해설 저작도구의 구현 예를 설명한다. 마지막으로, 4 절에서는 본 논문에 대한 결론을 맺는다.

2. 제안하는 화면해설 저작기술의 특징

가. 화면해설 삽입구간 추천

화면해설은 상황 변화적 요소나 자막, 그래픽 등 시각적 요소들을 대사나 중요 효과음이 없는 부분에 삽입하여 전체 프로그램의 이해를 저해하지 않도록 한다. 제안하는 화면해설 저작 방법에서는 효율적인 화면해설 작성을 위하여 방송스트림에 포함되어 있는 비디오, 오디오 및 자막 정보를 분석하여 화면해설 삽입 필요/가능 구간을 저작자에게 추천한다.

MPEG-2 TS 등 동영상 파일 내 비디오를 디코딩 한 후 화면전환검출(Shot Boundary Detection) 기술을 적용하여 검출된 화면 전환 시간 정보를 사용자에게 제시함으로써 장면(scene)이 바뀐 부분에 화면해설을 추가할 수 있도록 한다. 이를 위해 비디오 프레임 간 차분 히스토그램의 평균값을 이용하여 규정된 임계값(threshold)보다 클 경우 장면 전환이 일어난 것으로 판정한다.

오디오는 AC-3 디코더에서 출력된 스테레오 PCM 데이터를 입력 받아 비 대사(non-dialogue) 구간을 검출한다. 배경음악 및 효과음이 큰 경우에도 비 대사 구간 검출 성능을 보장하기 위하여 기존의 에너지 기반 VAD(Voice Activity Detection)[5]에 센터채널추출 기술을 접목하여 프레임 별 센터채널과 서라운드 신호 에너지 비율 변화를 추가 오디오 특징으로 사용하였다[6]. 이는 방송 콘텐츠 분석 결과, 대사 음성의 음상은 주로 센터채널에 존재하며 그 이외의 음향은 스테레오 음상 전역에 걸쳐 퍼져 있음에 기반한 비 대사 구간 검출 방법이다.

한편, 자막 방송을 위한 텍스트 데이터는 비디오 TS 패킷 내 비디오 ES (Elementary Stream) 의 Picture User Data 영역에 저장되어 전송된다. 따라서 해당 부분을 분석하여 자막 텍스트를 추출하고 관련 PES 헤더의 PTS (Presentation Time Stamp) 정보를 추출하여 비 대사 구간을 검출한다. 자막 텍스트 추출은 문장 단위로 수행하며 문장 단위는 문장 끝 부호(예로, .!? 등)를 검출함으로써 감지할 수 있다. 검출된 문장 끝 시점에서의 자막데이터 버퍼 내 첫번째 문자를 포함하는 TS 패킷의 PTS, 그리고 문장 끝 부호 캐릭터를 포함하는 TS 패킷의 PTS 를 알 수 있으므로 이를 기반으로 첫 캐릭터를 포함하는 TS 패킷의 상대 시간과 문장 끝 부호를 포함하는 TS 패킷의 상대 시간을 계산하여 비 대사 구간을 검출한다[7].

나. TTS 기반 화면해설 오디오 합성

TTS 란 글자, 문장, 숫자 등 텍스트를 사람의 음성으로 변환시켜주는 기술이다. TTS 는 시각 장애인을 위한 스크린 리더나 청각 장애인을 위한 대체 발성 수단 등 오랫동안 장애인 서비스 기술로써 사용되어 왔으며 근래에 와서는 음성합성기술의 발전으로 자연스러운 발음과 억양을 표현할 수 있게 되었다. 또한 부가적으로 말의 속도(speech rate), 크기(volume) 및 피치(pitch) 등 세밀한 제어가 가능하게 되었다.

제안하는 저작 방법에서는 화면해설 텍스트를 입력받아 TTS 를 이용하여 합성음성으로 변환함으로써 저작자가 화면해설 분량(글자 수 등)을 선정하는데 활용하거나 성우의 화면해설 더빙을 대체할 수 있다.

다. 화면해설 오디오 믹싱

동영상 파일 내 포함되어 있는 원 마스터 오디오와 화면해설 합성음성을 오디오 믹싱함으로써 최종적인 화면해설 오디오를 생성할 수 있다. 오디오의 자연스러운 흐름을 위해 화면해설 합성음성이 더해지는 구간의 마스터 오디오 볼륨을 조절할 수 있으며 해당 구간의 앞/뒤로 마스터 오디오에 페이드 인/아웃을 적용하여 믹싱할 수 있다.

3. 구현 예

제안한 화면해설 저작도구의 구현 예는 그림 1 과 같다.

입력된 동영상 파일(MPEG-2 TS, AVI, MP4 및 WMV 지원) 비디오는 Preview 창에, 오디오 파형은 Audio 창에 각각 보여진다. 저작자는 비디오/오디오/자막 자동 분석을 통해 화면해설 삽입구간 정보를 추천 받을 수 있으며 이 정보는 오디오 파형 밑의 창에 바(bar) 및 다이아몬드 형태와 색깔로 구분되어 보여진다. 이 정보를 활용하여 화면해설 에디터 창 내 화면해설 텍스트와 시간 정보를 추가하고 TTS 엔진을 실행하여 해당 시간에 합성음성을 출력한다. 구현 예에서는 디오텍(Diotek)의 Power TTS 를 이용하여 자연스러운 합성음성을 제공하였다[8].

마스터 오디오의 타임 라인에 맞추어 배치된 화면해설 합성음성과 마스터 오디오를 믹싱하여 화면해설 오디오를 wav 파일로 출력할 수 있으며 비디오와 함께 먹싱하여 화면해설 동영상 콘텐츠를 생성할 수 있다.

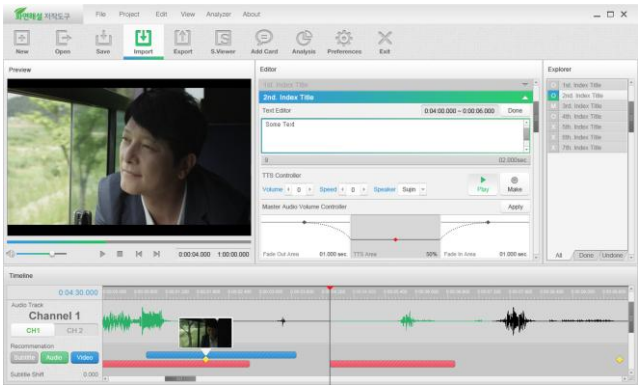

그림 1. 화면해설 저작도구 구현의 예

4. 결론

본 논문에서는 화면해설방송 저작 기술을 제안하고 각 세부 기술에 대해 설명하였다. 제안하는 방식은 화면해설 콘텐츠 저작 전반에 대한 편의성을 제공하며 특히, 화면해설 대본 작성에 효율적이다.

제안한 화면해설 저작도구를 기반으로 향후 실제 화면해설 콘텐츠 제작에서의 효율성을 검증하고 영상물뿐만 아니라 다른 서비스 영역으로의 확대를 위해 기능 확장 등 연구개발을 수행할 예정이다.

감사의 글

본 연구는 미래창조과학부가 지원한 2014 년 정보통신 · 방송(ICT) 연구개발사업의 연구결과로 수행되었음.

참고문헌

[1] 방송통신위원회고시 제 2011-53 호, “장애인방송 편성 및 제공 등 장애인 방송접근권 보장에 관한 고시”, 2011.12.26.

[2] Takayuki Ito, “ Activities to improve accessibility to broadcasting for visually impaired people,” IBM Workshop, Dec. 2010.

[3] M. Kobayashi and al, “ Unifying video captions and text-based audio descriptions,” CSUN 2011, San Diego, CA.

[4] Szarkowska, Agnieszka, “ Text-to-speech audio description: towards wider availability of AD,” Journal of Specialised Translation 15, 142-163, 2011.

[5] 임우택, 안충현, “TTS 를 이용한 화면해설 방송 제작 방법,” 한국방송공학회 하계학술대회, 2013.

[6] 장인선, 안충현, 장윤선, “ 화면해설방송 저작을 위한 비 대사 구간 검출,” 방송공학회논문지 제 19 권 제 3 호 296-306, 2014 년 5 월.

[7] 장인선, 임우택, 안충현, “ 화면해설방송을 위한 오디오/자막 기반의 무 대사 구간 검출,” 한국방송공학회 추계학술대회, 2013 년 11 월.

[8] http://speech.diotek.com/m20.php