다중 특징값을 이용한 교육용 어학 비디오의 내용기반 요약

한희준, 김천석, 추진호, 노용만
한국정보통신대학교 공학부 멀티미디어 그룹
hhj@icu.ac.kr

Content-Based Summarization of Educational Linguistic Video Using Multiple Features

Hee Jun Han, Cheon Seog Kim, Jin Ho Choo, and Yong Man Ro
Multimedia Group, Information and Communications University

요 약

방송 서비스상의 교육용 어학 컨텐츠의 증가와 더불어 비디오 컨텐츠의 효율적인 제공, 이용 및 관리를 위한 내용 기반 요약에 대한 연구가 필요하다. 본 논문에서는 교육용 어학 비디오의 내용 기반 요약을 위한 방법을 제안한다. 디지털 비디오로부터 샷 경계를 추출한 후 각 샷을 대표하는 키프레임으로부터 MPEG-7 비주얼 특징값들을 추출한다. 추출된 특징값들의 다중 조합을 통해 교육용 어학 비디오의 내용 정보를 세분화하여 요약 결과를 생성한다. 외국어 회화 컨텐츠에 대해 실험하여 알고리즘의 효용성을 검증하였으며, 제안한 방법은 교육용 방송 컨텐츠의 다양한 서비스 제공 및 관리를 위한 비디오 요약 시스템에 효율적으로 이용될 것이다.

1. 서론

현재 방송 서비스는 기존의 수동적인 단방향 방송에서 벗어나 소비자의 요구 사항을 만족시키는 양방향 방송 서비스로의 전환을 모색하고 있다. 양방향 디지털 방송 서비스는 소비자의 기호 및 성향에 적합한 프로그램, 요약된 방송 내용 및 하이라이트 장면등을 효율적으로 제공하는데 의의를 두고 있다. 즉, 시청자의 다양한 취향에 맞는 방송 컨텐츠를 제공하는 양방향 디지털 방송 서비스에 대한 요구가 증가하면서 방송 컨텐츠에 대한 요약, 검색 및 색인 기술 연구가 수행되고 있다. 더불어 방송 컨텐츠의 양이 증가함에 따라 서비스 제공자는 효율적인 컨텐츠 관리와 데이터 베이스 구축을 위해 비디오 요약 기술을 필요로 하게 되었다.

지금까지 비디오 컨텐츠로부터 비주얼 특징값들을 추출하여 내용기반으로 비디오를 요약하는 방법에 대한 여러 연구들이 이루어졌다. 특히, 칼라 및 에지, 움직임 정보 등을 이용해 뉴스 컨텐츠 및 스포츠 비디오에 대한 내용기반 요약에 대한 많은 연구가 실행되어 왔다[1-9]. 하지만 단일 특징만으로는 만족스러운 요약 결과를 얻지 못하며 비디오의 장르 및 스포츠 종목에 따라 적용해야 할 특징값들을 의존적으로 결정해야 하므로 일반적인 비디오 요약 방법을 지원하지 못한다는 단점을 지닌다. 또한 현재 교육용 방송 컨텐츠의 양은 급속하게 증가하고 있는 반면, 효율적인 서비스와 이용 및 컨텐츠 관리를 목적으로 하는 교육용 컨텐츠 요약에 대한 연구는 부족한 실정이다.

따라서 본 논문에서는 교육용 어학 비디오의 비주얼 특성을 파악하여 내용 기반 특징값들을 추출하고, 추출된 데이터를 다중 조합하여 어학 비디오의 세분화된 내용 정보를 생성한 후 의미있는 요약 결과를 도출하였다. 사용한 비주얼 특징들 중에서 일부는 국제 표준으로 정의된 MPEG-7 특징 정보를 사용하여 향후 재사용성 및 호환성을 고려하였다.

논문의 구성은 다음과 같다. 2절에서는 교육용 어학 비디오의 내용 분석 및 사용된 비주얼 특징들과 조합 방법, 요약에 필요한 비디오 내용 정보의 세분화 알고리즘을 포함하는 제안된 요약 방법에 대해 설명하고, 3절에서는 제안된 방법의 유효성 검증을 위해 MPEG-2 형식의 외국어 회화 교육용 어학 컨텐츠를 이용한 실험 결과를 기술한다. 마지막으로 4절에서는 결론에 대해 논한다.

2. 제안하는 방법

2.1 교육용 어학 비디오의 분석

교육용 어학 비디오 요약을 위해 의미있는 내용 정보를 분석할 필요성이 있다. 교육용 어학 비디오는 일반적으로 그림 1과 같이 사회자가 설명하는 부분(Explanation part), 외국인이 서로 대화하는 부분(Dialog part), 텍스트 정보로 이루어진 지문 부분(Text-based part)과 그밖의 유용하지 않은 정보를 담고 있는 잔여 부분(Remain part)으로 구성된다. 각 부분은 의미를 지닌 요소로 이루어지는데 그림 2와 같이 설명 부분은 일정한 장소(스튜디오)에 위치한 사회자가 어학 강의를 이룬다. 그리고 대화 부분은 사회자가 위치한 장소가 아닌 실외 및 실내의 다른 장소에서 외국인들끼리의 대화를 나타내며, 지문 부분은 비디오 화면에 텍스트 정보가 주류를 이룬다.

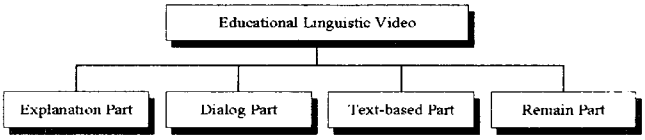

그림 1. 교육용 어학 비디오의 내용 세분화

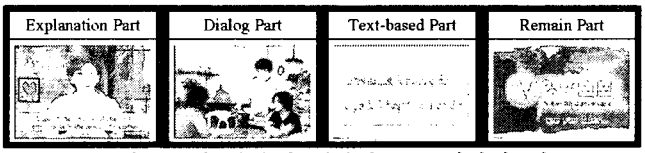

그림 2. 각 부분에 해당하는 프레임의 예

잔여 부분은 프로그램 타이틀이나 제작자 정보를 포함하고, 또는 화면 전환 부분을 나타내는데 이는 중요 교육 정보를 담고 있지 않을 뿐만 아니라, 사용자의 선호에 부합하는 교육 내용을 가지지 않는다. 그래서 교육용 어학 비디오 요약을 위한 생성 부분으로 간주하지 않는다. 따라서 교육용 어학 비디오로부터 설명 부분, 대화 부분, 지문 부분을 검출해 요약 정보를 생성하는 것을 목표로 한다.

2.2 요약을 위한 세부 알고리즘

본 절에서는 교육용 어학 비디오의 요약을 위한 세부 알고리즘에 대해 논한다. 그림 3은 비디오 요약의 세부 알고리즘이며 적용되는 특징값 및 적용 방법을 보여준다. 입력된 비디오로부터 먼저 샷 경계(Shot boundary)를 검출하고 각 샷을 대표하는 키프레임(Keyframe)을 추출한다. 추출된 키프레임으로부터 요약에 필요한 비주얼 특징값들을 추출한 후 미리 정의한 세분화된 내용

정보를 생성한다. 최종적으로 생성된 내용 정보는 MPEG-7 MDS 에 정의된 계층적 요약 (Hierarchical Summary) 구조에 맞추어 XML 문서 형식으로 표현된다[10] [11].

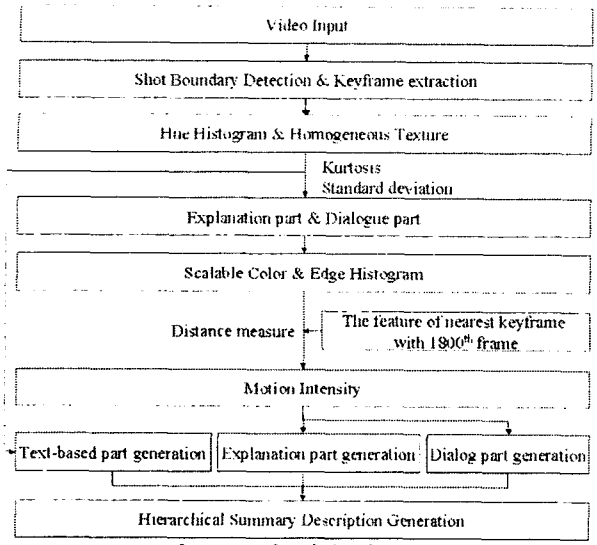

그림 3. 요약 세부 알고리즘

2.2.1 샷 경계 검출 및 키프레임 추출

교육용 어학 비디오의 세분화된 내용이 정의되면, 비주얼 특징 추출의 기본 단위인 샷과 키프레임이 추출된다. 본 알고리즘에서 사용한 샷 경계 검출 모듈은 MPEG-7 참조 소프트웨어 XM(experiment Model)에서 분리한 HierarchicalSummary DS 안의 샷 경계 검출 루틴을 개선하여 사용하였다[12][13].

샷 경계가 검출되면 각 샷을 이루는 프레임들 중에서 중간 프레임이 해당 샷을 대표하는 키프레임이 된다. 각 샷으로부터 키프레임을 구하는 루틴은 식 (1)과 같다. 여기서 *TotalShotNum* 은 검출된 샷의 개수, *Startframe* 은 샷의 시작 프레임을 나타내고 *Endframe* 는 샷의 끝 프레임을 의미한다.

$$for \quad n=1 \quad through \quad TotalShotNum \quad do \quad \{ \quad keyframe_n = (Startframe_n - Endframe_n)/2 \quad \} \tag{1}$$

교육용 어학 비디오를 L 이라 하고 s 는 샷, k 는 해당 샷을 대표하는 키프레임이라 하면 식 (2)와 같이 표현할 수 있다.

$$L=\{s_1, s_2, s_3, ..., s_i\}, \quad i = shot \ number \qquad k_i = keyframe \quad representing \quad s_i \tag{2}$$

2.2.2 적용되는 다중 특징값

2.2.2.1 색도 히스토그램 분포의 첨도
(Kurtosis of hue histogram distribution)

Hue histogram 은 이미지의 색도 분포를 나타낸다. 이미지의 RGB 값으로부터 Hue 값을 구한 후 360 빈마다의 분포 정도를 얻게 된다. 첨도는 분포의 뾰족한 정도를 나타내는 지표이며, 식 (3)은 각 샷을 대표하는 키프레임으로부터 색도 히스토그램 분포의 첨도를 구하는 방법이다.

$$k_i = \{h_i^1, h_i^2, h_i^3, ..., h_i^n\}, \quad n=360$$

$$Kur_i = \left\{ \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{p=1}^{n} \left(\frac{h_i^p - \bar{h}_i}{SD_i} \right)^4 \right\} - \frac{3(n-1)^2}{(n-2)(n-3)} \tag{3}$$

$$where \quad SD_i = \sqrt{\frac{n\sum_{q=1}^{n}(h_i^q)^2 - \left(\sum_{q=1}^{n} h_i^q\right)^2}{n^2}}$$

여기서 키프레임 k 는 Hue Histogram 데이터(h)들의 집합으로 나타낼 수 있다. *Kur* 은 Hue histogram 분포의 첨도값이며, *SD*는 h 로부터의 표준편차이다.

2.2.2.2 질감 기술자의 채널 에너지 표준편차
(Standard deviation of 30-channel energy)

이미지 질감 정보는 이미지의 균질성을 나타내는 패턴을 정의한다. 질감 특징 정보 표현을 위해 이미지의 주파수 영역을 6개 방향 성분과 5개 크기 성분으로 나누어 30개 채널에 대한 에너지를 구한다. 식 (3)은 질감 정보 추출을 위해 키프레임들의 30개 채널 에너지 표준편차를 구하는 식이다. e 는 채널 에너지를 나타내며 *SD _channelEnergy* 는 e 로부터의 표준편차이다.

$$k_i = \{e_i^1, e_i^2, e_i^3, ..., e_i^m\}, \quad m=30$$

$$SD_i\_ChannelEnergy = \sqrt{\frac{m\sum_{r=1}^{m}(e_i^r)^2 - \left(\sum_{r=1}^{m} e_i^r\right)^2}{m^2}} \tag{4}$$

2.2.2.3 에지 히스토그램 기술자
(Edge Histogram Descriptor)

이미지의 에지 정보를 나타내기 위해 먼저 이미지를 16개의 서브 블록으로 나눈 후, 각 블록에 대해 모두 5개의 에지 성분인 수직(vertical), 수평(horizontal), 45°, 135°, 무방향성(non-directional)을 기술한다. 각 블록의 에지 성분을 조합하여 한 이미지로부터 모두 80개의 local edge histogram을 얻으며, 16개 하위 블록의 local edge histogram을 조합하여 각각 5개의 golbal edge histogram과 40개의 semi-golbal edge histogram을 구성한다. 에지 히스토그램 기술자의 특징벡터 *ED* 는 식 (5)와 같다. 여기서 f_{local_s} 는 s 번째 local edge histogram빈을 나타내고, $f_{semi-golbal\_t}$ 는 t 번째 semi-golbal edge histogram빈을, f_{golbal_u} 는 u 번째 golbal edge histogram빈을 나타낸다.

에지 히스토그램 특징값들을 이용한 거리값 $dist_{ED}$ 는 각 키프레임들간의 유사도를 측정하는데 이용되고, 식 (6)과 같은 방법에 의해 구해진다. 식에서 *keyframe* 과 *keyframe'* 는 서로 다른 키프레임을 의미한다.

$$\overline{ED} = \begin{cases} f_{local\_1}, f_{local\_2}, ..., f_{local\_80} \\ f_{semi-global\_1}, f_{semi-global\_2}, ..., f_{semi-global\_40} \\ f_{global\_1}, f_{global\_2}, ..., f_{global\_5} \end{cases} \tag{5}$$

$$dist_{ED} = \sum_i \left| \overline{ED}_{keyframe}(i) - \overline{ED}_{keyframe'}(i) \right| \tag{6}$$

2.2.2.4 스케일러블 칼라 기술자
(Scalabel Color Descriptor)

이미지내의 칼라의 분포를 나타내는 히스토그램으로 표현된다. 이미지의 RGB 값들은 HSV 값으로 비선형 변환되며, HSV 칼라 공간을 모두 256개의 빈으로 나누고, 각 빈에 속하는 픽셀의 수를 측정함으로써 특징벡터를 구성한다. 칼라 기술자 특징값들을 이용한 거리값 $dist_{CD}$ 는 각 프레임들간의 유사도를 측정하는데 이용되고 식 (7)에 의해 계산된다.

$$dist_{CD} = \sum_i \left| \overline{CD}_{keyframe}(i) - \overline{CD}_{keyframe'}(i) \right| \tag{7}$$

2.2.2.5 움직임 강도 (Intensity of Motion Activity)

움직임 강도는 비디오 시퀀스내 객체의 움직임 정도를 일정 범위에 걸쳐 표현해 주는 특징값으로써, 움직임 강도(Motion Intensity)는 각 프레임의 16×16 매크로 블록으로부터 구해진 움직임 벡터들의 크기(Motion vector magnitude)를 프레임 해상도 (resolution)로 적절히 정규화하고 양자화시킨 값들의 표준편차이다.

다른 특징값 추출과는 달리 이것은 샷 단위로 연산되며,

비디오의 샷들로부터 움직임 강도를 구하는 방법은 식 (8)에 의한다. 여기서 mv_x 는 수평방향 움직임 벡터, mv_y 는 수직방향 움직임 벡터를 나타낸다. w 와 h 는 각각 프레임의 폭과 높이, n 은 샷에 속하는 프레임의 개수이며, mv_{mag} 는 움직임 벡터의 크기, Int_{motion} 는 mv_{mag} 의 표준편차로써 움직임 강도를 나타낸다.

$$mv_{mag} = \sqrt{mv_x^2 + mv_y^2}$$

$$Int_{motion} = \sqrt{\frac{\sum_{u=1}^{w\times h\times n}(mv_{mag})^2}{w\times h\times n} - \left(\frac{\sum_{u=1}^{w\times h\times n} mv_{mag}}{w\times h\times n}\right)^2} \quad (8)$$

2.2.3 비디오의 세분화된 내용 검출

그림 3에서 보는 바와 같이, 먼저 지문 부분을 생성하기 위하여 비디오로부터 색도 히스토그램(Hue Histogram)의 첨도와 MPEG-7 질감 기술자(Homogeneous Texture Descriptor)에 정의된 30채널 에너지 표준 편차를 조합한다. 지문 부분의 프레임들은 거의 동일한 색상의 배경을 가지므로, 색도 히스토그램 분포는 뾰족한 성향을 띄고 질감 기술자의 30채널 에너지 표준편차는 설명 부분이나 대화 부분에 비해 비교적 작은 값을 가지게 된다. 식 (3)에서 해당 샷을 대표하는 키프레임은 하나의 첨도값을 가지는데, 일정값 th_{Kur} 이상의 값을 가지는 키프레임은 지문 부분으로 결정한다. 또한, 각 샷을 대표하는 키프레임들은 각각 식 (4)의 채널 에너지 표준편차값을 한 개씩 가지는데, 이 값이 일정값 $th_{SD(Channel Energy)}$ 이하의 값을 가지게 되면 해당 키프레임을 지문 부분으로 결정하게 된다. 최종적으로 동시에 두 가지 임계값을 만족하는 키프레임들이 결정되고 그에 해당하는 샷들은 지문 부분의 세그먼트로 생성된다.

설명 부분과 대화 부분을 생성하기 위해서는 2.2.2.3~2.2.2.5 절에서 설명한 세 가지 특징값이 적용된다. 각 샷을 대표하는 키프레임들로부터 Scalable Color 와 Edge Histogram 기술자를 먼저 생성한다. 그리고 비디오의 1분 재생 후의 프레임(1800^{th} frame, 사회자의 설명 부분에 해당하는 프레임)과 가장 가까운 키프레임의 특징값과 식 (6), (7)에 의해 유사도 측정을 한다. 유사도 거리값이 일정값 th_{ED} , th_{CD} 이하 조건을 동시에 만족시키면 다시 식 (8)의 $Intensity_{motion}$ 특징값을 분석한다. 대화 부분은 다른 부분에 비하여 대화자들의 움직임 정도가 큰 반면, 설명 부분은 객체의 움직임이 거의 없는 샷들로 구성되기 때문에 $Intensity_{motion}$ 값이 일정값 이하이면 해당 키프레임을 설명 부분을 나타내는 샷이라 결정하고, 아니면 대화 부분을 나타내는 샷으로 결정하게 된다.

마지막으로 해당 부분으로의 샷 결정에 의해 생성된 세그먼트 정보를 이용하여 MPEG-7 MDS 에 정의된 계층적 요약 (Hierarchical Summary) 구조에 맞추어 XML 문서를 만들게 된다.

3. 실험 결과

제안한 교육용 어학 비디오 요약 방법의 효용성을 검증하기 위하여 EBS에서 제공하는 외국어 회화 컨텐츠를 이용하였다. MPEG-2 형식의 교육용 어학 비디오는 5 가지로 각각 중국어 회화, 영어 회화, 프랑스어 회화, 독일어 회화, 일본어 회화이며 각각은 20분 분량이다. 표 1은 실험에 사용된 5가지 비디오의 샷 경계 검출 및 키프레임 추출 결과로써 숫자는 샷의 개수를 나타낸다. 여기서 각 샷은 하나씩의 키프레임을 가진다.

표 1. 교육용 어학 비디오의 샷 검출 후 결과

	설명부분	대화부분	지문부분	잔여부분	전체
Chinese	55	39	2	14	110
English	40	41	0	11	92
French	50	34	1	14	99
German	37	39	3	10	89
Japanese	50	19	0	9	78

먼저 교육용 어학 비디오의 지문 부분을 검출하기 위하여 각 샷을 대표하는 추출된 키프레임들 각각으로부터 색도 히스그램 데이터를 구하여 첨도값를 얻었으며, 질감 기술자를 적용해 키프레임들의 30채널 에너지로부터 표준편차를 측정하였다. 그 다음 미리 정의된 임계값들을 적용하여 조건에 부합하는 특징값을 가지는 키프레임을 검출하였다. 검출된 키프레임은 각각 샷을 대표하므로 최종적으로 샷들은 지문 부분을 위한 세그먼트 정보를 이루게 된다. 표 2는 지문 부분을 생성하기 위해 비주얼 특징값을 적용한 결과이다. 중국어회화는 2개의 샷, 프랑스회화는 1개, 독일어회화는 3개의 샷이 지문 부분을 이루는데, 다중 특징값 조합에 따른 지문 부분 검출결과는 100%의 정확도를 보인다.

표 2. 지문 부분 검출결과

	total	correct	miss	false	recall	precision
Chinese	2	2	0	0	100%	100%
English	0	-	-	-	-	-
French	1	1	0	0	100%	100%
German	3	3	0	0	100%	100%
Japanese	0	-	-	-	-	-

실험에 사용한 비디오들은 설명 부분과 대화 부분을 비교적 많이 포함하고 있으며 따라서 해당 샷의 개수는 지문 부분에 비하여 많다. 설명 부분과 대화 부분을 위한 효율적인 요약 정보를 생성하기 위하여 적용할 Edge histogram, Scalable color, Motion intensity 세 가지 특징값들의 효율적인 조합이 중요하다.

표 3은 어학용 비디오로부터 설명 부분을 검출하기 위하여 Edge histogram 기술자에 정의된 특징값만을 이용한 결과이며, 표 4는 Scalable color 특징값을 이용한 설명 부분 검출결과이다. 그리고 표 5는 Edge histogram 과 Scalable color 특징값들을 동시에 적용하여 설명 부분을 검출한 결과이다. 보다 높은 정확도를 얻기 위하여 표 5로부터 얻은 결과에 다시 Motion intensity 특징값을 적용한다. 표 6은 최종적으로 앞에서 논한 세가지 특징값들을 효율적으로 적용하여 얻은 어학용 비디오의 설명 부분 검출결과를 보여주며, 높은 정확도를 나타낸다.

표 3. Edge histogram 특징값을 이용한 설명 부분 검출결과

	total	correct	miss	false	recall	precision
Chinese	55	48	7	14	87.27%	77.42%
English	40	40	0	49	100%	44.94%
French	50	50	0	11	100%	81.97%
German	37	37	0	42	100%	46.84%
Japanese	50	50	0	9	100%	84.75%

표 4. Sclalble color 특징값을 이용한 설명 부분 검출결과

	total	correct	miss	false	recall	precision
Chinese	55	53	2	50	96.36%	51.46%
English	40	39	1	2	97.50%	95.12%
French	50	49	1	25	98.00%	66.22%
German	37	34	3	6	91.89%	85.00%
Japanese	50	50	0	25	100%	66.67%

표 5. Edge histogram, Scalable color 특징값을 동시에 이용한 설명 부분 검출결과

	total	correct	miss	false	recall	precision
Chinese	55	48	7	13	87.27%	78.69%
English	40	39	1	2	97.50%	95.12%
French	50	49	1	8	98.00%	85.96%
German	37	34	3	6	91.89%	85.00%
Japanese	50	50	0	9	100%	84.75%

표 6. 설명 부분 검출결과

	total	correct	miss	false	recall	precision
Chinese	55	48	7	5	**87.27%**	**90.57%**
English	40	39	1	2	**97.50%**	**95.12%**
French	50	48	2	3	**96.00%**	**94.12%**
German	37	34	3	6	**91.89%**	**85.00%**
Japanese	50	49	1	2	**98.00%**	**96.08%**

설명 부분의 키프레임과의 Edge histogram, Scalable color 특징값 유사도 거리값이 임계값 이상이면 해당 키프레임들은 대화 부분을 구성하는 샷들을 대표하는 것들이다. 또한, 대화 부분을 구성하는 샷들은 움직임 정보가 많은 특성을 지닌다. 표 7은 Edge histogram, Scalable color, Motion intensity 비주얼 특징값들을 적용하여 얻은 대화 부분 검출결과이며, 그림 4는 결과의 예로써 일본어회화로부터 검출된 대화 부분을 보여준다.

표 7. 대화 부분 검출결과

	total	correct	miss	false	recall	precision
Chinese	39	34	5	7	87.18%	82.93%
English	41	40	1	1	97.56%	97.56%
French	34	31	3	2	91.18%	93.94%
German	39	35	4	3	89.74%	92.11%
Japanese	19	18	1	1	94.74%	94.74%

그림 4. 일본어회화의 대화 부분 검출결과

교육용 어학 비디오의 설명 부분, 대화 부분, 지문 부분이 검출되면, 각 부분에 대한 세그먼트 정보를 이용하여 MPEG-7 MDS 에 정의된 계층적 요약 (Hierarchical Summary) 구조에 맞추어 XML 문서를 생성한다.

그림 5는 영어 회화에 대한 요약 정보를 나타내는 XML 문서의 일부이며, 설명 부분, 대화 부분 및 지문 부분 검출결과 정보를 포함한 세그먼트 정보를 담고 있다. 이것은 방송 제공자가 시청자가 원하는 요약된 컨텐츠를 제공하는데 이용될 수 있다.

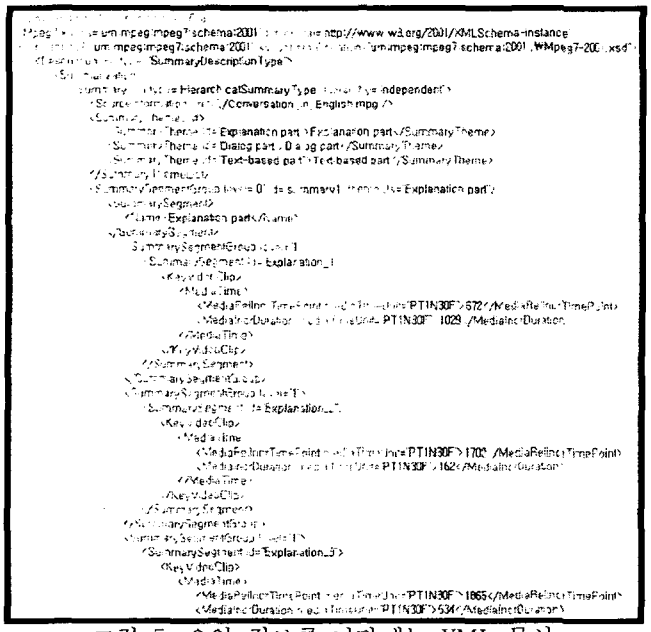

그림 5. 요약 정보를 나타내는 XML 문서

4. 결론

본 논문에서는 비디오의 비주얼 특성과 MPEG-7 국제 표준에서 제공하는 기술자들 일부를 이용하여 교육용 어학 비디오의 내용기반 특징값들을 추출하고, 이들을 조합하여 어학 비디오의 요약 정보를 생성하였다. 그리고 비디오의 구조적 내용 정보를 기술하는 요약문을 생성하였다.

교육용 방송 비디오는 요약된 정보를 제공할 필요성을 지닌다. 본 논문에서 제안한 방법은 교육용 어학 비디오에 대한 정확한 요약 결과을 제공할 것이며, 향후 양방향 방송 서비스상에서 소비자의 기호에 적합한 컨텐츠를 제공하는데 이용될 것이다. 제안한 방법에 의해 얻어진 최종 결과인 요약 정보를 표현하는 XML 문서는 시청자가 교육용 어학 컨텐츠의 내용을 신속하고 효율적으로 파악하는데 이용될 수 있다. 또한 시청자가 많은 교육용 어학 컨텐츠 중에서 선호하는 부분을 원할 때, 편의성을 제공하고 빠르고 정확한 검색도 제공할 수 있다.

향후 다양한 어학용 컨텐츠를 이용한 일반적인 요약 방법에 대한 연구가 필요하며, 실제 방송 환경을 고려한 요약 시스템 및 시청자의 선호도를 적용한 요약 방법에 대한 연구가 이루어져야 할 것이다.

Acknowledgement

본 논문은 정통부의 지원을 받아 수행중인 "DMB 기반 모바일 멀티미디어 응용기술 연구" 과제 수행 결과의 일부분이다.

참 고 문 헌

[1] Stefan Eickeler and Stefan Muller, "Content-Based Video Indexing of TV Broadcast News using Hidden Markov Models," ICASSP '99 Proceedings, IEEE, Vol. 6, pp. 2997-3000, March 1999.

[2] Ali M Dawood and Mohammed Ghanbari, "Scene Content Classification from MPEG Coded Bit Streams," Multimedia Signal Processing, IEEE, pp. 253-258, Sept. 1999.

[3] Dalong Li, Hanqing Lu, "Model Based Video Segmentation," SiPS 2000 IEEE, pp. 120-129, Oct. 2000.

[4] Hari Sundaram and Shih-Fu Chang, "Video Scene Segmentation Using Video and Audio Features," IEEE International Conference on Multimedia and Expo, pp. 1145-1148, 2000.

[5] Yan Liu and John R. Kender, "Video Frame Categorization Using Sort-Merge Feature Selection," Motion and Video Computing IEEE, pp. 72-77, Dec. 2002.

[6] Hee Kyung Lee, Cheon Seog Kim, Yong Ju Jung, Je Ho Nam, Kyeong Ok Kang, Yong Man Ro, "Video contents summary using the combination of multiple MPEG-7 metadata," SPIE Electronic Imaging, Vol. 4664, pp.1-12, 2002.

[7] Cheon Seog Kim and Yong Man Ro, "Semantic Event Detection using MPEG-7," SPIE Vol. 5021, pp. 372-379, 2003.

[8] Ichiro Ide, Koji Yamamoto and Hidehiko Tanaka, "Automatic Video Indexing Based on Shot Classification," AMCP'98, LNCS 1554, pp. 87-102, 1999.

[9] Nguyen Ngoc Thanh, Truong Cong Thang, Tae Meon Bae, Yong Man Ro, "Soccer Video Summarization System Based on Hidden Markov Model with Multiple MPEG-7 Descriptors," CISST 2003, Vol. 2, pp. 673-678, June 2003.

[10] Video Group, "Text of ISO/IEC 15938-1 FCD Information technology-Part 3 Visual," March 2001.

[11] Multimedia Descirption Schemes(MDS) Group, "Text of ISO/IEC 15938-5 FCD Information technology-Part 5 Description Schemes," March 2001.

[12] Toby Walker, Sanghoon Sull, "Proposal for a Video Summary Description Scheme," July 1999.

[13] Sang-Heun Shim, Seung-Ji Yang, Jeong-Hyun Yoon, Yong-Man Ro, "Real-time Shot Boundary Detection Based on Digital Video Cameras using the MPEG-7 Descriptor," 2001년도 한국 방송 공학회, p.193-198, 2001.