MPEG 시스템 스트림상에서 오디오 정보를 이용한 장면 경계 검출 방법

(A Scene Boundary Detection Scheme using Audio Information in MPEG System Stream)

김 재 홍 † 낭 종 호 †† 박 수 용 †††

(Jae Hong Kim)(Jong Ho Nang)(Soo Yong Park)

요 약 본 논문에서는 일반적인 영화를 인코딩한 MPEG 형식의 비디오 데이타에 대해 장면과 장면 사이의 경계점에서 나타나는 여러 오디오 특성을 이용하는 새로운 장면 경계 검출 방법을 제안하고 실험을 통해서 그 유용성을 보인다. 일반적인 영상에서 장면 경계 지점에서는 영상의 내용이 크게 바뀜에 따라 오디오 정보도 같이 변화한다는 특성이 있으며, 본 논문에서는 이러한 장면경계에서의 오디오 정보 변화를 각각 급진변화(Radical Change), 점진변화(Gradual Change), 미세변화(Micro Change)로 분류하였으며, 각 변화의 특성을 분석하고 이를 검출하는 알고리즘을 제안하였다. 급진변화는 장면과 장면의 경계점에서 오디오가 음량의 급격한 증감이 발생하고 음색 또한 급격히 달라지는 형태를 취하고 있으며, 점진변화는 긴 시간에 걸쳐서 음량 및 음색이 달라지는 형태를, 미세변화는 음량의 변화없이 일부 음색과 주파수 분포가 달라지는 특성을 가지고 있다. 본 논문에서는 이러한 특성을 토대로 시간축을 따라 진행하는 윈도우를 설정하여 이 윈도우 내에서의 오디오 변화를 추적함으로써 위의 세 가지 형태의 장면 경계를 추출하는 방법을 제안한다. 다양한 영화를 통한 실험에서 실제 샘플로 사용된 영화들에서 가장 많은 부분을 차지하는 급진변화에 대하여 본 논문에서 제안한 방법이 높은 검출율을 얻을 수 있음을 알 수 있었다. 본 논문에서 제안한 오디오 정보를 이용한 장면 경계 검출 방법은 비디오 정보를 이용한 장면 경계 검출과 같이 병행하여 사용함으로써 MPEG 형식의 영상정보에 대한 데이타 베이스 구축에 유용하게 사용될 수 있을 것이다.

Abstract This paper proposes a new scene boundary detection scheme for the MPEG System stream using MPEG Audio information and proves its usefulness by extensive experiments. A scene boundary has a characteristic that the audio as well as video information are changed rapidly. This paper first classifies this scene boundary into three cases ; Radical, Gradual, Micro Changes, with respect to the audio changes. The Radical change has a large-scale changing of decibel value and pitch value at a scene boundary, the Gradual change shows the long-time transition of decibel and pitch values from max to min or vice versa, and the Micro change displays a some change of pitch or frequency distribution without decibel changes. Upon this analysis, a new scene change detection algorithm detecting these three cases is proposed in which a progressive window with a time line is used to trace the changes in the audio information. Some experiments with various movies show that proposed algorithm could produce a high detection ratio for Radical change that is the most popular scene change in the movies, while producing a moderate detection ratio for Gradual and Micro changes. The proposed scene boundary detection scheme could be used to build a database for visual information like MPEG System stream.

· 본 연구는 1999년도 정보통신부 대학 기초연구 지원사업에 의한 것임

† 비 회 원 : 서강대학교 컴퓨터학과
jhkim@mljune.sogang.ac.kr
†† 종신회원 : 서강대학교 컴퓨터학과 교수
jhnang@ccs.sogang.ac.kr
††† 정 회 원 : 서강대학교 컴퓨터학과 교수
sypark@ccs.sogang.ac.kr

논문접수 : 1999년 9월 21일
심사완료 : 2000년 6월 3일

1. 서 론

최근 컴퓨터 하드웨어 및 압축기술의 발달로 인하여 여러 응용분야에서 멀티미디어 정보의 사용이 늘어나게 되었다. 이러한 변화에 덧붙여 네트워크의 고속화와 WWW의 인터넷 환경, 그리고 대용량의 저장 매체들의 등장은 종전에 찾아볼 수 없었던 VOD 서비스와 비디오 디지털 라이브러리 같은 첨단의 비디오 서비스 등을 가능하게 만들었다. 이러한 비디오 서비스를 가능하게 하기 위해서는 사용자들에게 다양한 형식의 검색을 지원하여야 하는데, 이는 비디오 내용을 기반으로 하는 자동화된 인덱싱 기술이 필요하게 됨을 의미한다.

비디오 데이타 인덱싱 기술에 관한 지금까지의 연구[12,13,14]는 대개 촬영자가 비디오 카메라를 동작시키고 멈춘 사이에 기록된 데이타인 샷(Shot)에 대한 인덱싱이 주를 이루었다. 하지만 이 샷에 의한 인덱싱은 그 분량이 많고 내용에 대한 의미 부여가 미흡하기 때문에 최근에는 비디오 데이타 중 의미있는 단위가 되는 장면(Scene)에 대한 인덱싱의 연구가 활발히 진행되고 있다. 장면 경계 검출에 대한 최근의 연구는 특정 장르를 가지는 비디오를 대상으로 하는 도메인 종속적인 방법[16], 장르 구분 없이 이루어지는 도메인 독립적인 방법[17], 그리고 샷의 특성에 의거한 유사성 비교에 의한 방법[15] 등이 있다. 이러한 일련의 연구들은 대부분이 비디오 영역에서만 이루어지고 있다. 하지만 장면의 경계에서 비디오 데이타의 변화가 없는 경우 장면 경계를 검출하기 어려운 경우가 있는데, 이런 경우 오디오 정보를 이용하여 장면 경계를 찾을 수가 있지만 현재 오디오에 의한 장면 경계 검출에 대한 연구는 거의 찾아 볼 수 없다. 다만 유사한 연구로써 오디오의 특성을 이용한 필름 장르의 검출[7]과 음성의 경계를 찾는 연구[1]가 있으며, 이외에 오디오를 대상으로 하고 있는 연구는 음성과 음악의 DB 구성에 필요한 인덱싱과 검출에 대한 연구[2,3,4,5]만이 있을 뿐이다.

본 논문에서는 일반적인 영화를 대상으로 한 MPEG 형식의 비디오 데이타에 대해 그 비디오 데이타가 가지고 있는 오디오 정보를 대상으로 장면과 장면사이의 경계점에서 나타나는 여러 특성을 이용하는 새로운 장면 경계 검출 방법을 제안한다. 일반적으로 영화에서 의미를 가지는 단위인 장면의 내부에서 오디오의 특성이 서로 비슷하게 나타나지만 서로 다른 장면과의 경계점에서는 오디오의 특성이 서로 달라지고, 또한 오디오의 하위 수준의 정보에서 변화가 많이 나타나는 성질을 가진다. 본 논문에서는 이러한 특성들을 이용하여 장면과 그 다음의 장면 사이에서 일어나는 오디오의 변화를 급진변화(Radical Change), 점진변화(Gradual Change), 그리고 미세변화(Micro Change)로 구분하고, 이들의 특성을 바탕으로 장면 경계를 찾아내는 방법을 제안하였다. 즉, 본 논문에서 제안하는 장면 경계 검출 방법은 일정 시간 동안 오디오의 주파수 영역의 하위 수준 데이타들을 분석하여 이전의 오디오 데이타들과 비교하여서 그 유사도를 측정하고, 그 유사도의 진행에 따라 앞에서 설명한 세 가지의 유형으로 분류하는 방법을 사용한다. 특히 장면 경계에서 가장 많이 나타나고 있는 경계 유형인 급진변화에 대하여 보다 정확한 검출 알고리즘을 적용하여 장면의 경계를 검출한다. 실험결과에 의하면 대상으로 하고 있는 영화들의 샘플 데이타에서 가장 많이 나타나는 유형인 급진변화에 대하여 높은 검출율을 나타내고 있다. 제안된 경계 검출 방법은 현재 많이 사용되고 있는 MPEG 형식의 비디오 데이타를 사용하는 데이타 베이스의 구축을 자동화하는데 유용하게 사용될 수가 있을 것이다.

본 논문은 다음과 같이 구성되어 있다. 2 장에서는 오디오를 이용한 인덱싱 관련 연구들과 경계 검출에 대한 유사 연구들에 대해 살펴보고, 3 장에서는 장면 경계에 대한 추상화, MPEG 오디오로부터 추출되어지는 하위 수준의 오디오 정보, 실제 장면 경계상에서의 하위 수준 정보들의 진행과 변화, 그리고 장르별 장면 경계의 분포에 대하여 알아보도록 한다. 4 장에서는 3 장에서 추상화되어진 장면 경계를 찾아내는 알고리즘에 대하여 설명하고, 5 장에서는 4 장에서 설명된 알고리즘을 적용하여 실제 장면 경계 검출을 수행한 실제 실험과 그 실험결과에 대하여 설명한다. 마지막으로 6 장에서 본 논문이 제시하고 있는 방법을 종합하여 정리하고 결론을 맺기로 한다.

2. 오디오 정보의 인덱싱에 관한 기존의 연구

멀티미디어 데이타를 인덱싱 하는 방법은 여러 가지가 있지만 크게 비디오 정보에 의한 인덱싱 방법과 오디오 정보를 통한 인덱싱 방법으로 나뉘어 진다. 현재 비디오 정보를 이용한 인덱싱 방법은 일반적으로 프레임(Frame) - 샷(Shot) - 장면(Scene)의 계층적 구조를 가지는 비디오 데이타의 특징상 이들의 경계 검출에 의한 인덱싱 방법[12,13,14,15]이 주를 이루고 있다. 이에 반해 오디오 정보를 이용한 인덱싱 방법은 현재 음성정보를 포함하고 있는 일반적인 영상매체를 대상으로 하고 있는 것은 최근에 Video Indexing의 보조 수단으로써 사용되고 있으나, 대부분 음악이나 영상이 없는 음

성 정보만을 대상으로 한 방법들이 대부분을 차지하고 있다. 본 장에서는 본 논문이 대상으로 하고 있는 영상정보인 MPEG 시스템 스트림에 포함되어 있는 MPEG Audio에 대한 개략적인 설명과, 영상정보의 장면 경계를 검출하기 위해 현재 사용되고 있는 독립적인 Audio (음악과 음성)들의 경계를 검출하는 방법들을 설명한다. 2.1절에서는 MPEG Audio에 대한 설명을, 2.2절과 3절에서는 오디오정보들을 대상으로 하고 있는 인덱싱 방법들과 오디오 경계 검출에 관한 기존의 유사한 연구들, 끝으로 2.4절에서는 영상정보에 담겨진 오디오 정보를 이용하여 영상정보의 인덱싱을 하는 관련연구에 대하여 알아보도록 한다.

2.1 MPEG Audio

MPEG Audio는 일반적인 고품질의 음성 정보를 낮은 전송률을 가지는 네트워크 상에서 재생이 가능하게 하기 위해 스트림 형식으로 압축을 행하고 있다. 이 MPEG Audio는 일반적인 음성정보의 저장형식인 PCM (Pulse Code Modulation)형식과는 달리 음성 정보를 주파수별로 분리를 하고 청각심리 모델링을 통하여 가청영역 밖의 데이타들을 버리고 주파수별 중요도에 따라서 할당 영역을 조절함으로써 음질 손상을 최소화하고 압축효과를 극대화하는 방법이다.

2.1.1 MPEG Audio의 인코딩 및 디코딩과정

MPEG Audio에서는 일반적인 PCM 데이타들이 단순히 Amplitude 값만을 저장하고 있는 것과는 달리 고품질의 음성 정보를 낮은 전송률 상에서도 유지하기 위해 압축작업을 행하고 있다. 이러한 압축 및 해제과정을 살펴보면 다음과 같다.

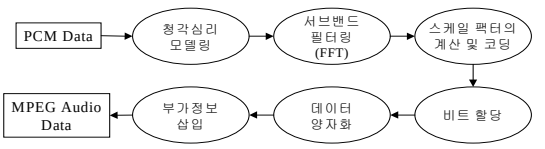


그림 1 MPEG Audio의 인코딩 과정

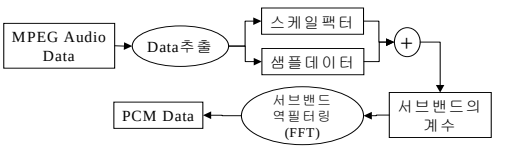


그림 2 MPEG Audio의 디코딩과정

위 그림에서 알 수 있듯이 MPEG Audio에서는 다음에 설명하고 있는 음성정보 인덱싱을 위해 사용하는 여러 가지 기법에서 사용하고 있는 요소들인 PCM계수들은 MPEG Audio 디코딩 과정에서 가장 최종 단계에서 나타난다. 이러한 기법들을 사용하기 위해서 MPEG Audio를 최종단계까지 디코딩을 행한다면 MPEG Audio의 디코딩과정에서 가장 많은 시간을 차지하는 역필터링 과정을 계속 거치게 되어 전체적인 시스템 성능 저하를 유발하게 된다. 따라서 본 논문에서는 이러한 성능저하를 막고 보다 빠른 작업을 수행하기 위하여 전체적인 디코딩 과정을 거치지 않고 FFT 과정 직전의 서브밴드 계수를 사용하고 있다. 이 서브밴드 계수들은 Amplitude 값을 나타내는 PCM 계수와 유사한 성질을 가지고 있고, 또한 32개의 밴드 영역을 가지고 있기 때문에 PCM 계수만으로는 분석하기 힘든 각 주파수별 특성을 분석하기가 용이한 장점을 가지고 있다.

2.2 음성 정보 인덱싱에 대한 연구

음성을 대상으로 하는 인덱싱은 그 목적이 대부분 화자의 전환, 성별의 감별 및 억양의 추출 등에 있다. 따라서 음성의 경계점을 찾는 것이 중요한 목표가 될 수 있다. <표 1>에서는 현재 음성정보의 인덱싱에 필요한 음성의 경계를 찾는 기존의 연구들을 정리하여 나타낸 것이다. <표 1>에 나타난 이러한 기법들은 단어, 화자, 성별의 변화요소들 같은 음성자체에 대한 분류에는 유용하게 이용될 수 있지만 현재 본 논문이 대상으로 하고 있는 일반적인 영상정보에 포함되어진 오디오 정보를 통한 영상정보의 인덱싱에는 보조적인 수단으로 사

표 1 음성의 경계를 찾는 관련연구

수행기법		측정대상	검출방법	관련연구
음소단위 분할 경계 검출		dB, ZCR	음성을 무성음과 유성음으로 나누어서 각각의 경계상에서 나타나는 dB차이와 Pitch에 해당하는 ZCR의 변화를 측정함으로써 음소단위의 경계를 검출	[1]
화자단위 분할 경계 검출	음의 에너지및 크기 측정	dB	구간내의 dB값의 통계적 분석에 따른 무성음 경계검출 및 묵음 검출	[2]
	부호변화 빈도측정	dB의 ZCR	소리가 가지는 Pitch 검출에 의한 성별 검출, 이성간 대화 전환	[3]
	유사도 측정기법	자기 상관 계수	소리가 가지는 여러 가지 요소들에 대한 유사성 검출에 의한 화자 전환 검출	[3]
	LPC 및 예측오류 측정	Sample 간 Difference	현재 샘플데이타의 통계적 수준에 따른 다음 샘플의 통계적 예측과의 차이를 검출함으로써 소리의 변화를 감지	[3]

용이 되고 있다. 또한 현재까지는 이러한 영상정보에 포함되어진 오디오 정보중 특히 MPEG Audio를 대상으로 한 연구는 아직 나오지 않고 있다.

2.3 비 음성 정보 인덱싱에 대한 연구

비 음성 정보는 크게 음악과 효과음으로 나뉘어 진다. 음악을 대상으로 한 연구들은 주로 악보를 기준으로 한 음표, 화음, 조, 마디, 리듬등에 대한 정보를 추출하여 각각의 곡들에 대한 인덱싱[4]을 수행하거나, 음악자체를 질의로 받아서 저장되어있는 음악을 검색하는 연구[5]가 있으며, 효과음에 대한 연구는 [6]에서 제안한 것처럼 효과음들이 가지는 몇몇 특징들을 가지고 이들을 분류하는 방법이 있다. 또한 이와 별개로 [7]에서 제안한 것으로 영상매체의 장르에 따라 나타나는 오디오의 특성들을 가지고 선택된 영상정보의 장르를 결정하는 방법이 있다. <표 2>는 이에 대한 분류를 정리하여 설명한 것이다.

표 2 비음성을 대상으로 한 기존의 연구

대 상		내 용	관련 연구
음악	음악자체의 인덱싱	음악을 스트로크와 패턴, 섹션이라는 계층적 구조로표현, 음의 기본음들의 연속성을 검출하고 분석하는 방법으로 곡에 들어있는 강약진행 및 사용악기, 음의 고저들에 대한 속성을 데이타 베이스화 한다.	[4]
	음악 검색을 위한 시스템	Humming의 형태의 사용자 입력을 Melodic Contour로 변형한 후 DB에 저장되어져 있는 MIDI형태의 곡들이 가지는 그것과의 비교를 통해 자동적인 검색 시스템 구축	[5]
효과음을 통한 범죄장면 검출		범죄장면에서 총성, 비명, 폭발음이 주를 이룬다는 점을 이용하여 효과음의 주파수 전이 특성을 미리 준비된 전형적인 자료들의 전이 특성과 비교하여 자동적으로 범죄장면을 검출하는 시스템	[6]
오디오 정보를 이용한 장르검출		각종 영상정보들이 장르별로 가지는 오디오의 특성을 통계적인 방법을 이용하여 자동적으로 영상정보의 장르를 결정하는 시스템	[7]

<표 2>에서 사용된 기법 중 효과음을 통한 범죄장면의 검출이나 오디오 정보를 통한 영상의 장르 결정부분은 영상과 관련한 오디오의 이용 방법들이지만 현재 본 논문에서 목표로 하는 영상정보의 경계 검출과는 약간의 거리를 두고 있다. 따라서 본 논문에서는 유사 연구들에서 사용된 기법들을 응용하여 음성만이 아닌 영상정보에 속해있는 오디오 정보를 대상으로 하여 그 경계를 검출함으로써 대상으로 하고 있는 영화들의 장면 경계를 검출하는 방법을 제안한다.

2.4 영상정보에 포함된 오디오 정보를 이용한 비디오 인덱싱에 관련된 연구

최근 들어 영상정보를 인덱싱하는 연구들 중 일부에서 오디오 정보를 보조적인 수단으로 사용하여 인덱싱을 하는 연구들이 진행되고 있다. 이러한 연구들은 대개 효과음이나 대화의 전환이나 단절을 이용하여 샷이나 장면의 경계를 찾아내는 방법들이다. 이러한 방법들은 <표 3>에 분류하여 설명하였다.

표 3 영상 정보에 포함된 오디오 정보를 이용한 비디오 인덱싱 관련 연구

대 상	내 용	관련 연구
오디오 정보 자체를 이용한 Segmentation	오디오의 Type(Speech, Silence등)에 관계없이 일정 길이의 2개의 Sliding Window간의 오디오 정보 차이를 검출하여 영상정보의 단절점을 찾아내는 시스템	[19]
오디오 정보를 영상 인덱싱의 보조수단으로 활용	오디오 정보를 음성인식을 통해 Text로 변환하여 영상정보와 같이 송출되어지는 Closed-Caption Text와 의 비교를 통해 영상의 장면변화를 검출하는 보조수단으로의 이용	[18]

3. 장면 경계에서의 오디오 변화 특성 분석

본 장에서는 MPEG 시스템 스트림 상에서의 장면 경계검출을 수행하기 위해서 사용된 방법들에 대하여 설명하도록 한다. 3.1절에서는 실제 영상정보의 장면 경계를 분류하고, MPEG오디오 정보로부터 추출 가능한 하위 수준 정보들에 대해 알아보고, 3.2절에서는 3.1절에서 추출되어진 하위 수준 정보들이 실제 장면 경계 상에서 보이는 변화들에 대하여 알아본다. 끝으로 3.3절에서는 영상 정보의 장르에 따른 장면 경계 분포를 알아봄으로써 장르에 따른 임계치 적용의 가능성에 대하여 알아보도록 하겠다.

3.1 장면 경계의 분류

일반적으로 장면의 변화가 있을 때에는 대개 내용의 변화가 있는 경우가 대부분이고 이때 나타나는 오디오의 데이타들도 같이 바뀌게 된다. 따라서 장면의 경계 지점에서는 앞쪽 장면의 오디오 정보와 뒤쪽 장면의 오디오 정보에 상당한 차이가 발생하게 된다. 본 논문에서는 이러한 차이를 <표 4>에서 나타나듯이 급진변화(Radical Change),점진변화(Gradual Change), 미세변화(Micro Change)의 세 가지로 구분을 하고 이에 맞추어

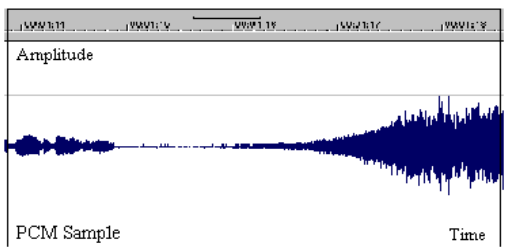
표 4 장면 경계의 분류

장면 경계	특 징	변화 대상	화면상 변화
급진 변화	장면 경계지점에서 짧은 시간 동안 급격한 오디오 정보의 변화가 일어난다.	음량 및 음색	내용상의 단절 및 상황의 급진전
점진 변화	장면 경계지점에서 긴시간동안 음량정보가 점진적으로 증감이 일어난다.	음량 정보	Fade In/Out 및 Dissolve
미세 변화	장면 경계지점에서 오디오 정보의 변화가 거의 일어나지 않는다.	일부 음색 정보	급진변화와 동일

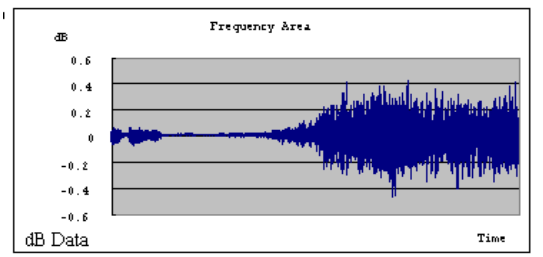
장면 경계를 검출한다. 이러한 경계들은 <표 6>의 장르별 장면 경계 분포에도 잘 나타나 있듯이 이 세 가지 변화가 영화진행상에서 가장 많은 빈도를 차지함을 알 수 있다.

위의 표에 나타난 장면 경계 지점들은 실제 영상에서도 비슷한 진행을 보이는 경우가 많기 때문에 비디오 정보들에 의한 장면 경계 검출 알고리즘과 병행하게 된다면 보다 정확한 장면 경계 검출을 수행할 수가 있을 것이다.

본 논문에서 사용하고 있는 하위 수준의 오디오 정보는 모두 주파수 영역내에 있는 dB값을 기초로 해서 추출되어 진다. 일반적인 오디오 정보는 Raw형태인 PCM



(a) PCM 샘플데이타



(b) (a)의 주파수 영역의 데이타

그림 3 PCM 과 주파수 영역 데이타간의 상관관계

(Pulse Code Modulation) 샘플데이타를 사용하지만 실제 주파수 영역내에 있는 dB값들도 이 PCM 데이타와 유사한 성질을 가지고 있다. 다음의 <그림 3>에서 (a)는 PCM 샘플 데이타를 나타내며, (b)는 이 PCM 데이타들이 FFT 변환과정으로 거친 후의 형태인 dB값들의 시간상의 진행과정을 나타내고 있다.(특히 (b)는 32개의 밴드중 가장 하위의 밴드값을 나타내지만 다른 밴드에서도 동일한 변화가 측정된다.) 이 그림에서 나타나듯이 PCM영역의 샘플데이타들이 FFT 변환을 거쳐 주파수 영역으로 변환이 된다 하더라도 데이타들의 기본 성향은 크게 달라지지 않음을 알 수가 있다. 현재 본 논문에서는 MPEG 오디오의 디코딩 과정에서 생성되는 이러한 FFT계수를 장면 경계 검출이 이용하고 있는데,

표 5 추출정보

추출 정보	의미 / 설명	검출 용도
프레임 dB평균	소리가 주파수 영역에서 가지는 값, FFT계수들의 프레임 평균 $F^{k}_{dB} = \sum_{i=0}^{N-1} \frac{\lvert x(i)\rvert}{N}$ x(i)는 FFT계수, k는 측정 윈도우 내의 프레임 번호, N 은 프레임의 크기, 프레임내의 dB의 평균으로써 인접 프레임간의 음량 변화를 측정하여 음의 단절여부를 검사한다.	음량 변화 검출
ZCR (영교차율)	FFT계수들이 일정시간동안 0을 지나는 횟수 $F^{k}_{ZCR} = \sum_{i=0}^{N-1} \frac{\lvert sgn(x(i)) - sgn(x(i+1))\rvert}{2}$ sgn(x)는 x의 부호가 양일 경우 1, 음일 경우 -1의 값을 가짐, 프레임내의 소리의 성격을 결정하는 수치로써 인접 프레임간의 이 수치가 차이를 보이면 다른 소리로써 판단하게 된다.	음색 변화 검출
Diff	한 프레임내에서 인접 샘플간의 차이 $F^{k}_{Diff} = \sum_{i=1}^{N-1} \frac{\lvert x(i) - x(i+1)\rvert}{N}$ 프레임 dB평균과 유사한 성질을 가지는 수치로써, 인접 프레임간의 음의 단절 여부를 결정하고, 그외에 일부 유사성 검출에도 사용된다.	음량 변화 검출
AMDF (평균차함수)	현재 프레임과 나머지 프레임의 샘플들의 차이 $AMDF(k) = \sum_{i=1}^{M} \lvert F^{k}_{dB} - F^{i}_{dB}\rvert$ M은 전체 윈도우의 크기. k는 현재의 프레임 번호, 이 수치는 윈도우내에 있는 전체 프레임간의 대표수치들간의 유사도를 측정하기 위해 사용된다.	유사성검출
자기상관계수	현재 프레임과 나머지 프레임간의 곱 $Autocorrelation(k) = \sum_{i=1}^{M} F^{k}_{dB} \cdot F^{i}_{dB}$ 이 수치도 앞서 설명한 AMDF와 같이 전체 윈도우내의 프레임간의 유사도를 측정하는 데 사용되고 있다.	유사성검출

이는 PCM 영역으로의 변환과정을 생략하게 되므로 전

체적인 속도 향상을 기대 할 수가 있다.

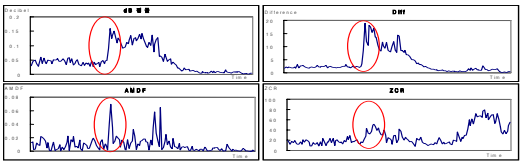
MPEG 오디오 스트림에서 복원되어진 밴드별 FFT 계수들은 장면의 경계점에서 여러 가지 특성을 띄게 되는데, 이러한 특성들을 보다 정확히 검출하기 위해서 본 논문에서 제안한 검출 방법에서는 FFT계수들로부터 <표 5>에 나타난 정보들을 추출한다.

다음의 <표 5>에서 제시된 추출 정보들은 이와 유사한 다른 연구[1,2,6,7]들에서도 사용되고 있기 때문에 이미 오디오 데이타들의 변화를 측정하는데 검증된 하위 수준 정보들이라고 할 수 있다. 본 논문에서는 경계 검출시 1초를 약 16개의 오디오 프레임으로 나누어서 각 프레임에 따로 MPEG 오디오 디코딩 시 생성되는 FFT 샘플 데이타들을 저장하고 또한 이 안에서 하위수준의 정보들을 샘플데이타들로부터 추출하여 약 5초의 길이를 가지는 윈도우 내에서 프레임단위의 실제 경계 검출을 수행하게 된다.(실제 실험에서 사용된 수치는, 1초의 각 밴드별 샘플개수가 44100/32 = 1379 그리고 16개의 프레임으로 나누었으므로 N=(44100/(32*16)=87, 그리고 5초의 윈도우크기를 가지므로 M=16*5=80이 된다.)

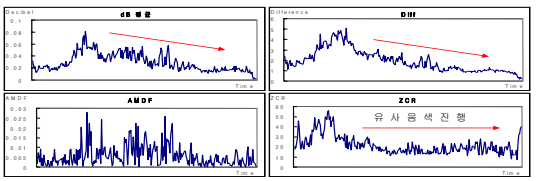
3.2 하위 수준 오디오정보 변화에 따른 장면 경계 분류

앞에서 설명한 3 종류의 장면 경계에서 <표 5>의 오디오에 대한 하위 수준 정보들이 어떻게 변하는가를 실험을 통하여 분석하였으며, <그림 4>는 이러한 분석의 예를 나타낸 것이다.

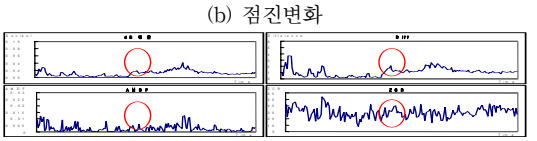
위 표에서 원형으로 표시된 지점이 실제 영상에서 장면의 경계가 나타난 지점이다. 위의 그림들이 보여 주듯이 분류되어진 세 가지의 장면 경계 유형들은 각각의 특징을 가지고 있다. 먼저 급진변화는 음량정보에 해당하는 dB과 Diff값이 짧은 시간동안 큰 폭의 변화를 보이는 특징과 유사도를 나타내는 AMDF와 ZCR 값 역시 큰 변화를 보이는 특징을 가진다. 점진변화는 음량정보인 dB과 Diff가 급진변화에 비해 긴 시간에 걸쳐 변화가 일어난다는 특징과, 이러한 음량변화가 일어나는 동안 유사도를 나타내는 ZCR값이 큰 변화 없이 일정하게 유지되는 특징을 가진다. 미세변화는 실제 장면의 경계가 나타났음에도 불구하고 오디오의 변화가 없기 때문에 위의 그림에도 나타나듯이 경계지점에서 측정 대상으로 하고 있는 하위 수준 오디오 정보에서 두드러진 변화가 없는 특징을 가진다. 따라서 설명한 세 가지의 경계 유형중 미세변화는 사실상 비디오 정보를 참조하지 않으면 검출되기가 어렵다.



(a) 급진변화



(b) 점진변화



(c) 미세변화

그림 4 장면 경계에 따른 실제 오디오 정보의 변화

3.3 장르에 따른 장면 경계 분포

오디오는 장르에 따라서 진행이 다양하게 이루어진다. 따라서 오디오의 장르를 살펴보게 되면 장면의 경계들이 어떠한 형태로 나타나는 가를 알 수 있게 된다. 위의 특성분석을 토대로 장르별 오디오 특성들을 알아보면 다음에 나오는 <표 6>과 같다. <표 6>은 대부분의 영화의 장면 경계지점에서는 음량과 유사도가 크게 변하는 급진변화가 대부분을 차지함을 보여준다. 본 논문에서는 별도의 장르별 특성에 따른 검출을 하지 않지만 전체적으로 가장 많이 나타나는 장면 경계인 급진변화의 검출률이 다른 변화들의 비해 높게 나타나고 있다.

표 6 장르별 장면 경계분포

장르	특징	발생빈도가 높은 장면 경계유형
Action	1. 상황전개가 빠르고 장면간의 경계가 분명 2. 후반부로 갈수록 급진변화의 비율이 높아짐 3. Burst 발생비율이 높다	급진변화 (68%)
Drama	1. 대화에 의한 진행이 대부분 2. 하나의 장면의 길이가 타 장르에 비해 길다	미세변화 (66%)
Musical	1. 내용의 변화가 음악의 변화와 같이 나타남 2. 타장르보다 점진변화가 많음	급진변화(71%)와 점진변화(21%)
미스테리 / 추리	1. 급반전 상황변화가 많음 2. 대화에 의한 상황변화가 많음	급진변화(63%)와 미세변화(30%)

4. 장면 경계검출 알고리즘

위에서 살펴본 바와 같이 영상에서의 장면의 경계점에서는 대개 내용 자체가 변하는 경우가 많기 때문에 그러한 내용 변화에 맞추어 오디오 정보가 같이 변한다는 것을 알 수가 있다. 본 논문에서는 장면 경계상에서의 하위정보들의 변화를 이용하여 실제 장면 경계를 검출하는 알고리즘을 제안하도록 한다.

4.1 알고리즘 설계

본 논문에서는 1/16초간의 하위수준정보들을 가지는 프레임들을 약 5~10 초의 시간을 가지는 윈도우 내에 버퍼링을 하고 이 윈도우 내에서의 변화를 추적하는 방법을 사용한다.

4.1.1 급진변화의 추적

급진변화 변화는 앞서 설명한 것처럼 짧은 시간에 정보들이 급격히 변화하기 때문에 비교적 쉽게 검출이 가능하다. 급진변화 변화는 dB값과 음색의 변화에 많이 나타나므로 기본적으로 이 값들의 추적하게 되고, 나머지 값들은 보조적인 역할을 담당하게 된다. 일단 검출의 기본대상이 되는 조건은 윈도우 내에 있는 Frame들 사이의 dB값을 비교하는 방법을 사용하며, 이를 식으로 나타내면 다음과 같다.

$$Detected = \frac{\max(|F^{i}_{dB}|)}{\min(|F^{i}_{dB}|)} > Threshold \quad (1)$$

$$DetectedTime = Time(\min(|F^{i}_{dB}|)) \quad (2)$$

$\max(|F^{i}_{dB}|)$는 윈도우내의 최대 음량값,

$\min(|F^{i}_{dB}|)$는 윈도우내의 최소 음량값,

$Time(\min(|F^{i}_{dB}|))$은 최소 음량이 위치하고 있는 시간상의 위치를 의미한다.

즉 검출 윈도우 내에서 max dB값과 min dB값의 비율이 설정된 임계치 이상 된다면 이러한 상황이 장면 경계 검출의 대상이 된다. 하지만 실제적으로는 이러한 조건을 만족하는 것은 장면 경계뿐만 아니라 다른 상황에서도 많이 나타날 수가 있는데, 대표적인 것은 장면 내부에서 폭발이나 비명 같은 Burst가 발생하는 경우, 전체적인 음량이 작게 나타나는 경우, 그리고 장면의 내부에서 상황의 급작스러운 전개가 일어나는 경우 등이다. Burst의 발생 경우는 전체적으로 소리가 작을 때 일시적으로 소리가 커지게 되거나, 그 반대의 경우가 있고, 전체적인 음량이 작게 나타나는 경우는 dB값들이 작게 나타나기 때문에 별로 크지 않는 변화에도 위의 식을 만족하는 검출 오류(false positive)의 가능성이 있다. 그리고 장면의 내부에서 상황의 급작스러운 반전 등이 일어날 때에는 내용상의 분위기가 바뀌기 때문에 전체적인 배경음악이나 효과음 등이 큰 변화가 오게 된다. Burst 경우는 시간적인 필터링과 윈도우간의 전체적인 정보들의 비교 및 분석을 통해서 어느 정도 제거가 가능하고, 전체적인 dB값이 작게 나타나는 경우에는 검출의 대상이 되는 $\max(|F^{i}_{dB}|)$와 $\min(|F^{i}_{dB}|)$값의 차이에 따른 차등적인 임계치 적용에 의해 어느 정도 제거가 가능하지만, 상황의 급작스러운 반전에 의한 검출은 비디오 화면이 없는 상황에서는 제거가 불가능하다. Burst 가 나타난 경우를 제거하는 시간적인 필터링 방법은 아래와 같다.

$$Time(Detected(i)) - Time(Detected(i-1)) < BurstThreshold \quad (3)$$

Detected(i) : i 번째 검출지점

검출이 되어진 두 지점간의 시간차가 설정되어진 시간 임계치 보다 크다면, 이는 실제로도 다른 장면일 가능성이 높기 때문에 앞서 검출된 지점은 장면의 경계가 된다. 하지만 시간차가 시간 임계치 보다 작다면 일단 일시적인 Burst의 가능성이 높기 때문에 다시 검출이 일어난 두 윈도우간의 유사도를 측정하게 된다. 여기서는 dB의 AMDF값이나 ZCR값을 측정함으로써 유사도를 측정하고 있다.

$$AMDF(Detected(i)) - AMDF(Detected(i-1)) < AMDFThreshold \quad (4)$$

$$ZCR(Detected(i)) - ZCR(Detected(i-1)) < ZCRThreshold \quad (5)$$

$AMDF(Detected(i))$: 현재 검출 지점의 AMDF 값

$ZCR(Detected(i))$: 현재 검출 지점의 ZCR값

위의 식에서 유사도를 나타내는 AMDF와 ZCR의 값이 주어진 임계치 보다 작다면, 검출되어진 두 경계 지점은 동일한 장면으로 인식되어 뒤에 검출되어진 지점이 제거가 되며, 임계치를 넘어서게 되면 두 경계 지점은 완전히 서로 다른 장면으로 구별이 되어 모두 장면의 경계로서 검출이 된다. 그리고 dB값이 작은 상태에서 발생하는 검출오류를 제거하기 위한 차등 임계치 적용은 $\min(|x(i)|)$의 값을 크기에 따라 여러 단계로 나누어 작은 값을 가지는 단계에 속할수록 검출 오류를 보이기 쉽기 때문에 보다 큰 임계치를 적용하여 검출의 정확도를 높이는 방법이다.

4.1.2 점진변화의 추적

기본적으로 점진변화의 검출은 급진변화와 크게 다르지 않다. 하지만 점진변화의 경우 Fade In/Out의 경우는 급진변화 검출과 유사한 방법으로 검출이 가능하지만 Overlap이 되는 경우에는 실제 윈도우 내에서 max/min값이 크게 나타나지 않기 때문에 검출이 용이하지 않다. 또한 긴시간에 걸쳐서 일어나기 때문에 진행시간이 윈도우 사이즈를 넘어설 경우에는 검출이 되지 않을 수도 있다. 기본적인 검출 순서는 다음의 식으로 나타낼 수 있다.

$$Detected = \frac{\max(|F^{i}_{dB}|)}{\min(|F^{i}_{dB}|)} > Threshold \quad (6)$$

일단 기본적인 검출방법은 급진변화 검출과 동일하게 시작되어진다. 하지만 급진 변화는 짧은 시간동안에 변화를 하지만 점진변화는 긴 시간에 걸쳐서 일어나기 때문에 Max 지점과 Min지점 사이에 긴 시간적 차이가 나타나고 그 사이에 있는 정보들이 시간 축에 따라 일정한 증가 혹은 감소하는 성질을 나타내기 때문에 급진변화 검출과는 다른 조건들이 뒤따르게 된다. 다음의 식들은 그러한 조건들을 나타내고 있다.

$$|Time(\max(|F\ dB_{i}|)) - Time(\min(|F^{i}_{dB}|))| > TimeThreshold \quad (7)$$

$$|dB(i)| \cong \min(|F^{i}_{dB}|) + \frac{\max(|F^{i}_{dB}|) - \min(|F^{i}_{dB}|)}{Time(\max(|F^{i}_{dB}|)) - Time(\min(|F^{i}_{dB}))} \quad (8)$$

$$Time(\min(|F^{i}_{dB}|)) < i < Time(\max(|F^{i}_{dB}|)) \quad (9)$$

즉, max와 min 값들이 검출된 시간의 차이가 급진변화와는 달리 상당히 긴 시간이 되며, 그 사이 dB값의 분포가 일차함수 적인 증감을 나타내는 성질을 가진다. 이때 시간차가 설정된 시간 임계치 보다 작게되면 이는 점진변화가 아닌 급진 변화로써 인식이 된다. 또한 최대/최소값 시간 사이의 임의의 시간의 하위 수준 정보들의 값이 시간에 비례하여 일차 함수적인 증가 및 감소를 보이지 않는다면 이 역시 급진 변화로 인식이 된다. 앞쪽 장면과 뒤쪽 장면의 소리가 서로 Overlap이 되는 경우에는 검출이 용이하지 않지만 일부 몇 가지의 경우 성격이 서로 다른 소리들이 겹쳐지게 된다면 음색의 유사성을 검출하는 방법으로 일부 경우를 검출할 수 있다. 이 경우 제일 먼저 유사성 검출의 정보가 되는 ZCR과 프레임간의 AMDF값 그리고 자기상관계수들을 dB값들보다 우선하여 검색을 한다. 이 ZCR의 최대 최소값들을 분석함으로써 음색의 유사성을 통해 장면의 경계를 유추하는 것이다. 하지만 대개 급진 변화의 경우 ZCR값도 동시에 급격한 변화를 보이는 것이 많기 때문에 이 ZCR값의 검출 후 바로 dB값의 비교 분석을 수행함으로써 급진변화의 경우를 제거한다.

$$Detected = \frac{\max(F^{i}_{ZCR})}{\min(F^{i}_{ZCR})} > Threshold \quad (10)$$

그리고

$$\frac{\max(|F^{i}_{dB}|)}{\min(|F^{i}_{dB}|)} > Threshold \quad (11)$$

4.1.3 미세변화의 추적

미세변화는 대개 영화에서 침묵이 지속되는 가운데 장면의 변화가 있는 경우나 아무런 배경 음악이나 효과음 없이 대화만이 지속되는 가운데 장면의 변화가 오는 경우가 이에 해당되어 진다. 일반적으로 대화의 경우 녹음 시 음량이 그리 크지 않기 때문에 전체적인 dB에 미치는 영향이 그리 크지 않기 때문에 dB값들이 아주 약하게 나타난다. 따라서 dB값에 의한 검출은 사실상 불가능하게 된다. 하지만 일부의 경우 이 미세변화를 일부 검출할 수 있는 경우가 있는데, 이 경우는 대화 전환에 의한 장면의 변화시 화자의 성별이 바뀌는 경우 일부 검출이 되어 진다. 이 경우 유사도를 나타내는 ZCR값 변화를 추적하여 프레임간 ZCR값들의 유사도를 측정하여 장면의 경계를 검출하게 된다. 아래의 식은 이러한 조건들을 나타낸 것이다.

$$Detected = \frac{\max(Autocorrelation_{ZCR}(k))}{\min(Autocorrelation_{ZCR}(k))} > Threshold \quad (12)$$

$$Autocorrelation_{ZCR}(k) = \sum_{i=1}^{N} (F^{k}_{ZCR})(F^{i}_{ZCR}) \quad (13)$$

4.2 전체적인 알고리즘

앞에서 장면 경계의 유형별 특징들을 알아보고, 그에 따른 검출 방법에 대하여 설명하였다. 하지만 실제적인 구현에 있어서는 위의 경우를 따로 따로 검출하는 것이 아닌 통합적인 검출이 되어야 한다. <그림 5>에서 실제적인 장면 경계 검출 알고리즘을 나타낸 것이다.

4.3 변화유형과 장르에 따른 다른 임계치 적용

위의 알고리즘에서 적용되는 임계치들은 변화 유형과 장르에 따라 보다 정확한 검출을 수행하기 위해서 서로 다른 임계치를 적용하였다. 각 장르들은 자장면이 우세하게 가지는 변화에 따라서 이 임계치의 적용을 받게 된다. 변화별 임계치 적용은 다음의 <표 7>과 같다. 아래에서 사용된 임계치들은 사용된 샘플들의 각 요소별 수치들의 통계적인 분포에 따라서 적용한 값들이다. 이 임계치들은 5.2.3절에서 설명하듯이 Recall/Precision 값이 많은 실험을 통하여 어느 한쪽에 치우치지 않고 동시에 높은 값을 만족하는 임계치로써 설정되었다. 특히 dB 값의 경우 오디오의 녹음 상태에 따라서 그 dB 값이 어느정도 차이가 나게 되고, 특히 dB 수치 자체가 전체적으로 높은 경우와 낮은 경우 동일한 비율이 적용될 수가 없기 때문에, 최소 dB값에 따라서 차등적인 비율을 적용하였다.

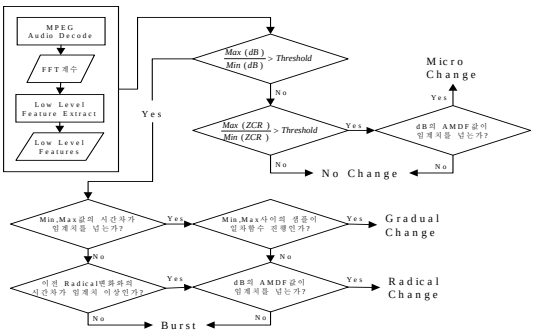


그림 5 전체 수행 알고리즘

표 7 변화별 적용 임계치

유형	대 상	조 건	임계치
급진 변화	dB 비율	최소 dB < 0.005	10
		0.005 ≤ 최소 dB < 0.075	7
		0.075 ≤ 최소 dB < 0.01	5
		0.01 ≤ 최소 dB	3
	시간차	이전 급진변화와의 시간차	2초
	AMDF(평균차함수)	급진변화 발생지점의 dB의 AMDF값	0.05
점진 변화	dB 비율	급진변화와 동일	
	시간차	이전 급진변화와의 시간차	2초이상
	AMDF	사용하지 않음	
미세 변화	ZCR	최대 ZCR / 최소 ZCR	5
	AMDF	ZCR검출 지점의 ZCR AMDF값	0.1

5. 실험 및 결과분석

5.1 장면 경계 검출 시스템 구현

본 장에서는 4 장에서 제안하였던 장면 경계 검출 알고리즘의 유용성을 보이기 위해 실제적으로 MPEG 오디오 스트림을 입력받아 실제 경계 검출을 수행하는 프로그램을 구현하였다. 여기서 입력으로 받는 MPEG 오디오 스트림들은 실제 영화가 담겨있는 MPEG 시스템 스트림으로부터 오디오 부분을 추출해낸 것을 사용하였으며, MPEG 오디오 Layer 2로 압축되어 있다. 실제 구현은 펜티엄 2 프로세서와 윈도우즈 98을 운영체제로 하고 있는 PC에서 MS Visual C++ 5.0을 통해 이루어졌다.

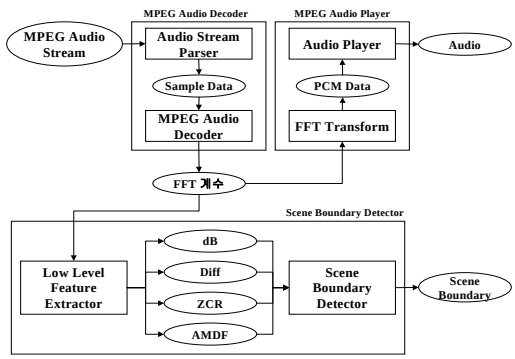


그림 6 시스템 구성도

전체적인 시스템의 구성도는 <그림 6>과 같이 크게 MPEG 오디오 스트림을 디코딩하는 부분과 디코딩과정에서 생성되는 FFT 계수들을 넘겨받아 이를 토대로 실제 장면의 경계점을 찾는 부분과 오디오 재생을 하는 부분으로 나뉘어 진다. Audio Stream Parser는 Header와 Sample Data로 구성되어 있는 MPEG 오디오 스트림에서 헤더정보와 실제 Sample Data들을 추출하는 하여 MPEG Audio 디코더로 보내주는 역할을 하고 있다. MPEG Audio 디코더는 스트림 파서로부터 넘겨받은 헤더 정보와 압축상태의 샘플데이타들을 가지고 실제 디코딩을 수행하여 FFT 계수들을 추출하여 오디오 재생부분이나 장면 경계 검출 부분으로 넘겨주는 역할을 담당하고 있다. 이렇게 추출되어진 FFT 계수들은 FFT 변환과정을 통해 PCM 샘플데이타로 변환된 후 오디오 재생에 사용되거나 장면 경계 검출을 위한 하위수준정보들을 추출하기 위해 이용되어 진다. 실제 장면 경계 검출을 진행하는 부분은 FFT 계수들로부터 하위수준의 정보들을 추출한 후 이들의 진행유형에 따라서 미리 모델링된 세 가지의 장면 경계 형태와 비교하여 장면 경계 검출을 수행한다.

구현된 프로그램은 MPEG Simulation Group[11]에서 공개용으로 제공되어진 MPEG 오디오 재생기인 maplay를 기반으로 하여 만들어 졌다. MPEG 오디오 스트림을 디코딩하고 오디오 재생을 하는 부분은 maplay의 내부 모듈을 사용하고 있으며 디코딩 중간과정에 FFT 계수 추출부분과 하위수준정보 추출부분, 그리고 장면 경계 검출 부분을 별도로 제작하여 삽입한 형태를 띄고 있다. 현재 구현된 프로그램은 디코딩과정에서 나오는 총 32 개의 밴드중 주로 가장 뚜렷한 변화를 보이는 배경음악이나 효과음들의 저음영역이 위치하게 되는 최하층 밴드 영역을 사용하여 4장에서 제안한

알고리즘을 적용, 장면 경계 검출을 수행하고 있다.

5.2 실험 및 분석

5.2.1 실험결과

본 논문에서 구현한 장면 경계 검출 알고리즘의 실제 실험에서 각각의 하위 수준 정보들의 비교에 사용되어진 임계값들은 실험을 통하여 결정하였다. 실험에 사용되어진 데이타들은 총 5 편의 영화를 MPEG 시스템 스트림으로 만들어 놓은 것을 오디오 부분만을 추출하여 사용하고 있으며, 모두 Layer-2 의 압축형식을 가지고 있다. 실제 실험에 의한 아래의 <표 8>에 나타나 있으며, 이후의 <표 9>은 급진 변화가 적은 드라마 장르를 제외한 실험 결과이다.

위에서 나타난 결과를 놓고 볼 때 변화별 유형으로는 전체적으로 가장 많은 빈도수를 보이는 급진 변화에 대

표 8 실험결과

Title	유형	실제 경계	검출	Cor-rect	Incor-rect	Miss	Recall	Preci-sion
Starship Troopers (60Min) Action장르	**Radical**	**19**	**33**	**18**	**15**	**1**	**0.947**	**0.545**
	Gradual	4	10	1	9	3	0.250	0.100
	Micro	5	8	1	7	4	0.200	0.125
	Total	28	51	20	31	8	0.714	0.392
Jackie Brown (60Min) 드라마장르	**Radical**	**7**	**13**	**4**	**9**	**3**	**0.571**	**0.308**
	Gradual	2	2	1	1	1	0.500	0.500
	Micro	19	32	8	24	11	0.421	0.250
	Total	28	47	13	34	15	0.464	0.277
Little Mermaid (76Min) 뮤지컬장르	**Radical**	**24**	**48**	**20**	**28**	**4**	**0.833**	**0.417**
	Gradual	7	6	2	4	5	0.286	0.333
	Micro	3	5	0	5	3	0.000	0.000
	Total	34	59	22	37	12	0.647	0.373
X-File (60Min) 미스테리/ 추리장르	**Radical**	**25**	**38**	**20**	**18**	**5**	**0.800**	**0.526**
	Gradual	4	5	1	4	3	0.250	0.200
	Micro	13	15	3	12	10	0.231	0.200
	Total	42	58	24	34	18	0.571	0.414
Snake Eyes (68Min) 미스테리/ 추리장르	**Radical**	**25**	**42**	**19**	**23**	**6**	**0.760**	**0.452**
	Gradual	2	1	1	0	1	0.500	1.000
	Micro	11	15	3	12	8	0.273	0.200
	Total	38	58	23	35	15	0.605	0.397
Total	**Radical**	**100**	**174**	**72**	**102**	**28**	**0.720**	**0.414**
	Gradual	19	24	6	18	13	0.316	0.250
	Micro	51	75	15	60	36	0.294	0.200
	Total	170	273	93	180	77	0.547	0.341

표 9 드라마 장르를 제외한 실험결과

Title	유형	실제 경계	검출	Cor-rect	Incor-rect	Miss	Recall	Preci-sion
Without Drama Genre	**Radical**	**93**	**161**	**77**	**84**	**16**	**0.828**	**0.478**
	Gradual	17	22	5	17	12	0.294	0.227
	Micro	32	43	7	36	25	0.219	0.163
	Total	142	226	89	137	53	0.627	0.394

한 검출률이 높게 나타나고, 장르별 검출률은 급진변화의 비율이 높은 장르일수록 전체적인 검출율이 높게 나타남을 알 수가 있다. 여기서 사용된 Precision과 Recall 수치는 잘못 찾은 경우와 찾지 못한 경우에 대한 제대로 찾은 경우의 비율을 의미한다.

5.2.2 오류 발생의 유형별 분석

위의 실험결과에서 나타난 잘못 검출되는 경우와 찾지 못하는 경우는 다양하게 나타난다. 이러한 오류가 검출될 수 있는 경우를 각 유형별로 알아보면 다음의 <표 10>과 같다.

<표 10>에 나타난 결과들을 놓고 볼 때 오디오 정보의 변화들은 실제 장면에서 벌어지는 사건들의 영향을 많이 받고 있음을 알 수가 있다. 실제 장면 경계 지점이 아닌 부분인 상황 급변 지점에서 급진 변화로 인식하는 오류를 범하고 있는 것이 이 사실을 뒷받침 해준다. 이런 경우에는 오디오 정보만이 아닌 비디오 정보의 지원을 받아야만 올바른 경계 지점을 인식할 수가 있을 것이다.

5.2.3 임계치 변화에 따른 Recall과 Precision의 변화

위의 실험에서 사용된 임계치는 샘플들의 통계적 분포에 맞추어 최적화 된 임계치들이다 이러한 임계치를 변화시켰을 때 실제적인 검출 Precision과 Recall의 수치는 다음과 같이 변화하게 된다. 여기서는 앞서 설명한 <표 7>에 나타나있는 음량정보의 임계치 만을 변화시켜서 급진변화와 점진변화의 검출에 대하여 실험을 하였다.

실제 적용되는 임계치들을 변경한 결과 임계치를 높게 적용한 경우 전체적인 검출율(Recall/Precision)이 모두 감소하였고, 임계치를 낮게 적용한 경우 전체적인 Recall 수치는 높아 졌지만 Precision이 낮게 나타남을 알 수가 있다. 임계치를 높게 적용했을 때 Recall/Precision이 동시에 낮아지는 이유는 임계치를 높게 설정하게 되면 보다 비율이 크게 나타나는 변화만이 경계 검출의 대상이 되어 실제 검출의 대상이 줄어들게되어 실제 경계가 Missing되는 경우도 같이 발생하게 되므로 Recall수치가 감소하게 되고, Precision의 경우에는 실제 검출대상이 줄어들게되어 Precision값은 유지되거나

오히려 증가되어야하나, 실제 Reject(틀린 것을 검출하지 않는 경우) 대상보다 Missing의 대상이 더 많아지는 상황이 발생하여 Precision값이 줄어드는 경향을 나타내게 된다. 임계치를 낮게 적용했을 경우에는 검출대상이 많아 지게 되어 찾지 못했던 경우들도 찾아내게 되어 전체적인 Recall 값은 증가하게 되지만, 오히려 False Positive의 경우가 Correct의 경우보다 많아 지게 되어 전체적인 Precision 값이 감소하게 된다.

표 10 장면 경계 유형에 따른 오류 유형

장면 경계 유형	오류유형	원인
급진 변화	Incorrect	장면 내부에서 갑작스러운 상황변화가 발생하는 경우 전체적인 오디오의 변화가 발생한다 (102/174).
	Missing	연속적인 Burst발생과 Burst에 의한 장면 경계 발생의 경우 찾지 못하는 경우가 발생한다 (28/100).
점진 변화	Missing1	검색 윈도우 보다 큰 시간적 길이를 가지는 점진변화 발생 시 찾지 못하는 경우가 발생한다 (6/19).
	Missing2	점진변화가 진행 중에 Burst가 발생할 경우 급진변화로써 인식한다. 하지만 이 경우는 제대로 찾은 경우로 포함하였다(3/19).
	Missing3	Overlap이 발생하는 경우 Max/Min의 비율이 임계값 이하로 발생하여 검출이 되지 않는 경우가 발생한다(4/19).
미세 변화	Incorrect	동일 장면 내에서 단순한 화자 전환에도 유사도가 임계값보다 낮게 나타나는 경우 장면 전환으로써 검출이 되는 오류가 발생한다(60/75).
	Missing	음량의 변화가 없을 경우 음의 유사도가 크게 차이 나지 않는 경우 검출이 되지 않는다 (36/51).

표 11 임계치를 높게 적용한 경우(음량 임계치 - 15,10,7,5)

Title	유형	실제 경계	검출	Cor-rect	Incor-rect	Miss	Recall	Preci-sion
Total	**Radical**	**100**	**155**	**55**	**100**	**45**	**0.550**	**0.355**
	Gradual	19	20	3	17	16	0.158	0.150
	Micro	51	75	15	60	36	0.294	0.200
	Total	170	250	73	177	97	0.429	0.292

표 12 임계치를 낮게 적용한 경우(음량 임계치-7,5,3,2)

Title	유형	실제 경계	검출	Cor-rect	Incor-rect	Miss	Recall	Preci-sion
Total	Radical	100	192	76	116	24	0.760	0.396
	Gradual	19	26	6	20	13	0.316	0.231
	Micro	51	75	15	60	36	0.294	0.200
	Total	170	293	97	196	73	0.571	0.331

5.3 비디오 정보에 의한 장면 경계 검출과의 비교

비디오 정보를 이용하여 영상정보의 장면 경계를 검출하는 방법은 현재는 그리 많지 않다. [15] 에서는 이미 검출되어진 샷(Shot)들의 키 프레임(Key Frame)을 추출하여 이들 Key Frame간의 유사도를 측정함으로써 비슷한 유사도를 가지는 샷끼리 묶어 줌으로써 장면을 추출하는 일련의 과정을 설명하고 있다. [15]에서는 비교적 샷간의 유사도가 높은 드라마와 드라마 식 구성을 가지는 영화를 대상으로 실험을 하여 약 80% 이상의 Recall 과 75% 이상의 Precision을 보이고 있다. 하지만 드라마식 구성이 아닌 액션물이나 미스테리/추리등의 장르를 가지는 영화에서는 오디오에 의한 장면 경계 검출이 비교적 우세한 경향을 나타낸다. 따라서 [15]에서 구현되어진 비디오 정보에 의한 장면 경계 검출과 본 논문에서 제시하고 있는 오디오에 의한 장면 경계 검출이 결합하게 된다면 서로의 약점을 보완하는 관계가 되는데 그 이유는 비디오 정보에 의한 장면 검출은 크게 두 가지 경우에서 각각 Missing과 False Positive를 나타내게 된다. 첫 번째는 동일 장소에서 갑작스러운 상황변경으로 인한 전혀 새로운 전개가 나타나는 경우인데, 이 경우에는 동일장소와 동일 인물들이 나타나기 때문에 비디오 정보로써는 그 경계를 인식을 할 수가 없다. 하지만 오디오 정보에서는 상황이 바뀌었기 때문에 급박한 오디오정보의 변화가 일어나게 되어 그 장면의 경계를 검출할 수 있게 된다. 두 번째는 비디오 내부에서 반복적인 대화장면이 나타나는 경우 비디오 정보가 계속해서 크게 변하는 경우가 많기 때문에 장면의 경계로써 잘못 인식할 확률이 매우 크게 나타나지만 오디오의 경우 큰 변화를 보이지 않기 때문에 경계가 아닌 것으로 정확히 인식이 가능해진다.

6. 결론 및 앞으로의 연구방향

일반적인 비디오 데이타들을 이용하여 구성이 되는 디지털 라이브러리나 VOD시스템들을 구축하기 위해서는 사용자들의 요구에 상응하는 비디오 정보에 대한 별도의 DB 구축이 필요로 하게 된다. 이러한 별도의 DB

구축을 위해서는 비디오 데이타에 대한 자동적인 내용 기반의 인덱싱 작업이 필요하게 되는데, 일반적으로 비디오 정보에 대한 인덱싱은 주로 샷이나 장면 단위의 인덱싱으로 주를 이룬다. 하지만 이러한 일련의 인덱싱 작업에 대한 연구는 음성 정보를 제외한 영상정보를 이용하여 이루어져 왔다. 본 논문에서는 MPEG 오디오 스트림을 대상으로 오디오 정보를 이용한 새로운 장면 경계 알고리즘을 제안하고, 제안되어진 알고리즘을 적용한 장면 경계 검출기를 구현하였다.

본 논문에서 제안한 장면 경계 검출 방법은 비디오에서 일반적으로 장면 경계 지점에서 오디오 정보들이 이전 장면과는 다른 형태의 정보유형을 가진 다는 점을 이용하고 있다. 이러한 변화를 검출하기 위해서 본 논문에서는 장면의 경계 상에서 하위수준 오디오 정보들이 보이는 변화 유형에 따라서 세 가지의 장면 경계 유형을 제시하였고, 또한 이러한 유형들에 맞추어 MPEG 오디오 디코딩 과정에서 생성되는 FFT계수들로부터 여러 가지 하위수준정보들을 추출하여 이들의 변화를 세 가지 장면 경계 유형에 따라 분류하여 각각의 경계를 검출하는 방법을 제시하였다. 또한 모든 하위 수준의 정보들을 MPEG오디오 디코딩과정에서 생성되는 FFT계수를 이용함으로써 PCM형태의 오디오 정보를 주파수 영역으로 변환하기 위한 FFT변환에 따른 오버헤드를 제거하여 전체적인 속도 향상을 꾀하였다. 실험결과에서 나타나듯이 실제적으로 대부분의 영화에서는 급진변화가 주종을 이루고 있고 또한 본 논문에서 제시한 알고리즘도 이에 맞추어 급진변화 검출에 높은 정확도를 보이고 있다. 이에 반해 점진변화나 미세변화에 대해서는 그다지 높은 검출율을 보이지는 못하고 있다. 결국 전체적인 검출율은 평균적으로 비디오가 오디오에 비해 높게 나타나기 때문에, 오디오 정보만을 이용한 장면 경계 검출은 어느 정도 한계를 가지게 된다. 하지만 비디오 정보만으로는 결코 찾을 수 없는 부분들이 다수 존재하고 있고, 이러한 부분을 오디오 정보를 이용한 방법이 해결해줄 수 있기 때문에 비디오와 오디오 정보는 상호 보완적인 관계를 가지게 된다.

앞으로의 연구는 높은 검출율을 보이지 못한 점진변화와 미세변화에 대해 오디오의 다른 정보들을 이용하여 보다 검출율을 높이는 것과, 비디오 정보를 이용한 경계 검출방법에 오디오 정보를 이용한 검출 방법을 추가하여 현재 보다 정확히 장면의 경계를 찾는데 주력할 것이다.

참 고 문 헌

[1] O. Essa, "Using Prosody in Automatic Segmentation of Speech," *Proceedings of the 36th Annual Conference on Southeast Conference,* 1998.

[2] R. Gonzalez, and K. Melih, "Content Based Retrieval of Audio," *Proceedings of the Australian Telecommunication Networks & Applications Conference (ATNAC'96)*, 1996.

[3] B.S. Atal, and L.R. Rabiner, "A Pattern Recognition Approach to Voiced-Unvoiced-Silence Classification with Applications to Speech Recognition," *IEEE Transactions on ASSP, Vol.24, No.3*, 1976.

[4] P. Aigrain, P.J.V. Longueville, and P. Lepain, "Representation - based user Interfaces for the Audio Visual Library of Year 2000," *Proceedings of SPIE Multimedia Computing and Networks 1995*, Vol.2417, 1995.

[5] A. Ghias, J. Logan, D. Chamberlin, and B.C. Smith, "Query by Humming : Musical Information Retrieval in an Audio Database," *Proceedings of ACM Multimedia'95*, 1995.

[6] S. Pfeiffer, S. Fischer, and W. Effelsberg, "Automatic Audio Content Analysis," *Proceedings of the fourth ACM International Multimedia Conference*, 1996.

[7] S. Fischer, R. Lienhart, and W. Effelsberg, "Automatic Recognition of Film Genre," *Proceedings of the third International Conference on Multimedia '95* , 1995.

[8] ISO/IEC/JTC1/SC29/WG11, *Coding of Moving Pictures and associated Audio for Ddigital Storage Media at up to about 1.5Mbit/s - Part3 : Audio, ISO/IEC International Standard 11172-3*, Aug. 1993.

[9] 정제창, *그림으로 보는 최신 MPEG*, 교보문고, Dec. 1995.

[10] K.R. Rao, and J.J. Hwang, *Techniques & Standards for Image, Video & Audio Coding*, Prentice Hall Press, 1996.

[11] http://www.mpeg.org/, MPEG Simulation Group.

[12] H.C. Zhang, and G.L. Zick, "Scene Decomposition of MPEG Compressed Video," *Digital Video Compression : Algorithm and Technologies*, SPIE Vol.2419, 1995.

[13] J. Meng, Y. Juan and S.F. Chang, "Scene Change Detection in a MPEG Compressed Video Sequence," *Digital Video Compression : Algorithm and Technologies*, SPIE Vol.2419, 1995.

[14] J.R. Kender and B.L. Yeo, "Video Scene Segmentation via Continuous Video Coherence," *Proceedings of Computer Vision and Pattern Recognition*, 1998.

[15] 이숙경, *MPEG 비디오 스트림에서 줄거리 특성에 기*

초한 신 경계 검출 방법, 서강대학교 석사 학위 논문, 1998.

[16] H.J. Zhang, Y. Gong, and S.W. Smoliar, "Automatic Parsing of News Video," *Proceedings of IEEE Conference on Multimedia Computing and Systems,* 1994.

[17] M. Yeung, B.L. Yeo, and B. Lui, "Extracting Story Units from Long Programs for Video Browsing and Navigation," *Proceedings of International Conference on Multimedia Computing and Systems,* 1996.

[18] A.G. Haupmann, and M.J. Witbrock, "Story Segmentation and Detection of Commercials in Broadcast News Video," *Proceedings of the Advances in Digital Libraries Conference,* 1998.

[19] J.S. Boreczky and L.D. Wilcox, "A Hidden Markov Model Framework for Video Segmentation Using Audio and Image Features," *Proceedings of ICASSP,* 1998.

김 재 홍

1997년 8월 서강대학교 전자계산학과 졸업(학사). 1999년 8월 서강대학교 대학원 컴퓨터학과 졸업(석사). 1999년 9월 ~ 현재 서강대학교 대학원 컴퓨터학과 박사과정. 관심분야는 멀티미디어 시스템. 멀티미디어 스트리밍 기술, Audio 관련 기술, 인터넷 프로그래밍 등

낭 종 호

1986년 2월 서강대학교 전자계산학과 졸업(학사), 1988년 2월 학국과학기술원 전산과 졸업(석사), 1992년 2월 한국과학기술원 전산과 졸업(박사). 1992년 3월 ~ 1992년 8월 한국과학기술원 정보전자연구소(연구원), 1992년 9월 ~ 1993년 8월 일본 Fujitsu 연구소(방문연구원). 1993년 9월 ~ 현재 서강대학교 컴퓨터학과 부교수. 관심분야는 멀티미디어 시스템, 병렬 프로그램 언어 및 시스템, 인터넷 프로그래밍 및 응용

박 수 용

정보과학회논문지 : 소프트웨어 및 응용 제 27 권 제 1 호 참조