

다중 오디오 특징을 이용한 유해 동영상의 판별

Classification of Pornographic Video with using the Features of Multiple Audio

김정수, Jungsoo Kim*, 정명범, MyungBum Chung**, 정보경, BoKyung Sung**,
권진만, JinMan Kwon**, 구광효, KwangHyo Koo**, 고일주, Ilju Ko***

요약 본 논문에서는 인터넷의 역기능으로 현대 사회에 큰 문제를 야기 시키는 음란성 유해 동영상을 내용기반으로 판별하기 위한 방법을 제안하였다. 유해 동영상에서 오디오 데이터를 이용하여 특징을 추출하였다. 사용된 오디오 특징은 주파수 스펙트럼, 자기상관, MFCC이다. 음란성의 내용이 될 수 있는 소리의 특징을 추출하였고 동영상 전체 오디오에서 해당 소리의 특징과 일치하는지를 측정하여 유해성을 판별하였다. 제안한 방법의 실험은 각 특징마다 유해 판별 측정 결과와 다중 특징을 이용한 측정 결과를 비교 수행하였다. 하나의 오디오 특징만을 추출하여 사용하였을 때 보다 다중 특징의 사용이 좋은 결과를 얻을 수 있었다.

Abstract This paper proposed the content-based method of classifying filthy Pornographic video, which causes a big problem of modern society as the reverse function of internet. Audio data was used to extract the features from Pornographic video. There are frequency spectrum, autocorrelation, and MFCC as the feature of audio used in this paper. The sound that could be filthy contents was extracted, and the Pornographic was classified by measuring how much percentage of relevant sound was corresponding with the whole audio of video. For the experiment on the proposed method, The efficiency of classifying Pornographic was measured on each feature, and the measured result and comparison with using multi features were performed. I can obtain the better result than when only one feature of audio was extracted, and used.

핵심어: *Classification of Pornographic, Multiple Audio Features, Frequency Spectrum, Autocorrelation, Mel frequency cepstral coefficients*

*주저자 : 숭실대학교 미디어학과 e-mail: dotline@ssu.ac.kr

**공동저자 : 숭실대학교 미디어학과

***교신저자 : 숭실대학교 미디어학과 교수; e-mail: andy@ssu.ac.kr

1. 서론

컴퓨터와 인터넷의 발달로 디지털 콘텐츠가 급속도로 확산되고 있다. 최근 인터넷을 통한 동영상 콘텐츠의 규모가 커지고 있으며 그 역기능으로 유해성 유해 동영상에 의한 피해도 날이 갈수록 커지고 있다. UCC의 발달로 영상 콘텐츠의 대규모 업로드와 인터넷 공유 사이트 등이 그러하며 이를 통해 무분별하게 유해 콘텐츠가 퍼지고 있다. 실제로 대형 포털 사이트에서 유해 동영상이 장시간 노출 되어 큰 파장이 일어난 적이 있다. 이에 따라 청소년들은 컴퓨터를 통해 쉽게 유해 동영상 콘텐츠에 노출 될 수 있는 문제가 발생 되었다. 이러한 유해적 요소를 담고 있는 영상 콘텐츠는 다양한 사회적 사건, 사고를 일으킬 수 있으며 특히 청소년들에게 더 크게 부정적인 영향을 미칠 수 있다.[1]

이같이 청소년들에게 피해를 주는 유해 동영상을 사전에 미리 차단해 주기 위해서 컴퓨터가 내용적으로 비교하여 판별하는 시스템이 필요하다.[2] 내용기반 시스템은 멀티미디어 데이터 정보의 내용(contents)을 수학적으로 분석하여 처리하는 시스템이다. 유해 판별을 위한 내용기반 시스템은 대체로 동영상의 이미지 정보를 이용한 비전 처리 분야가 많이 연구되었다.[3][4] 하지만 단일 매체만을 이용한 유해 판별은 정보량이 적기 때문에 분석 결론을 유추하기에 부족하다. 따라서 이미지 정보와 함께 오디오 정보 또한 유해를 판별할 수 있는 요소로 사용되어야 한다. 본 논문에서는 오디오 정보를 이용한 내용기반 유해 동영상 판별 시스템을 제안한다. 이를 위해 유해 동영상에서 나타나는 음란성의 내용이 될 수 있는 소리의 특징을 추출하고 이를 이용해 유해를 판별한다.

2. 오디오 특징 추출

유해 동영상에서 오디오를 이용한 유해를 판별하기 위해 동영상에서 오디오 데이터를 가져오고, 해당 오디오에서 성행위 시 나타나는 유해성 오디오 데이터의 특징들을 추출한다. 특징은 주파수 스펙트럼(Frequency Spectrum), 자기 상관(Autocorrelation), 멜 주파수 캡스트럼 계수(MFCC)를 사용하며, 각 특징들은 유해성 판별을 하기 위한 요소로 사용된다.

2.1 주파수 스펙트럼 (Frequency Spectrum)

유해성 오디오의 신호와 무해한 오디오의 신호는 다른 주파수 대역을 가질 것이다. 따라서 주파수 스펙트럼을 통해 유해성 오디오 신호에 포함되어 있는 주파수 성분에 대한 크기를 특징으로 사용한다. 본 논문에서는 프라트 음성 분석 프로그램을 이용하였다.[5] 프라트는 암스테르담 대학의

Paul Boersma와 David Weenink가 제작한 음성 분석 프로그램으로 인터넷 홈페이지[8] 에서 무료로 다운로드 받을 수 있다. 이 프로그램을 이용하여 시계열 오디오 데이터에서 주파수 스펙트럼으로 변형하여 특징을 추출 한다.

그림 1은 프라트 분석 프로그램을 이용하여 시계열 오디오의 데이터를 주파수 스펙트럼으로 나타낸 결과를 보여준다. 유해성 오디오에서는 대부분 초기 스펙트럼 그래프에 굴곡이 2번 나타나며, 스펙트럼의 주파수 대역이 비교적 넓지 않은 특징을 보인다.

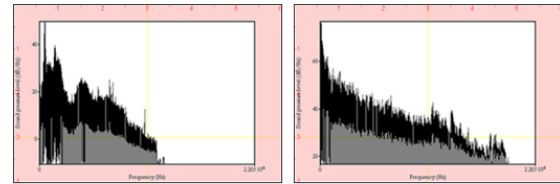


그림 1. 주파수 스펙트럼 특징 그래프

2.2 자기 상관 (Autocorrelation)

유해성 오디오의 신호는 불규칙 적이지 않고 반복적인 패턴을 보인다. 이러한 반복적인 패턴을 측정하기 위해 자기 상관(Autocorrelation)의 값을 계산한다.[6]

자기 상관은 시계열 데이터에서 현재의 상태가 미래의 상태와 연관 갖는 것을 의미한다. 즉, 오디오의 시계열 데이터 x_1, \dots, x_i 에서 $(x_1, x_{1+k}), \dots, (x_i, x_{i+k})$ 로 k 간격마다 데이터의 조합을 만들었을 때, 조합된 두 데이터 사이의 각조에 공통된 계통적 관련이 있을 때 이 시계열은 k차의 자기상관을 갖는다. 이러한 k차 자기 상관을 통해 현재와 미래의 상태의 유사관계를 측정하며, 조합된 두 데이터 사이의 관계를 측정하기 위해 차의 절대 값을 사용한다.

k차의 자기 상관 유사관계 측정은 현재의 상태와 단 한순간의 미래의 상태를 비교하지 않고, 매번 미래의 상태와 비교한다. 이는 k를 한 차수만을 이용하는 것이 아니라 $1, \dots, n$ 차를 사용 하는 것을 의미한다. 시계열 데이터가 가지는 모든 미래의 상태에 따른 유사관계를 측정함으로써 미래의 유사관계 변이를 알기 위함이다. 유사관계 변이는 반복적인 패턴을 보이는 유해성 오디오에서는 낮게 나타나며 팝음악 같이 음이 다채로운 오디오에서는 높게 나타난다.

그림 2는 $1, \dots, n$ 차에 따른 상관관계를 그래프로 나타낸 것이다. 그래프에서 차이 값이 낮을수록 상관관계가 높은 것이며, k에 따른 값의 편차가 크지 않으면 미래에 따른 상관관계의 변이가 낮음을 의미한다. 따라서 유해성 오디오의 자기 상관은 상관관계의 차이 값이 낮으며, 변이 또한 낮은 특징을 보인다.

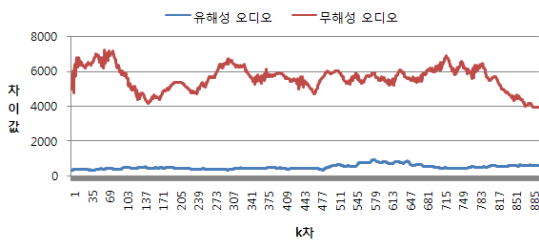


그림 2. 자기 상관 특징 그래프

2.3 멜 주파수 캡스트럼 계수 (MFCC)

MFCC[7]는 인간의 청각 특성을 모델링하는 방법으로 음성을 인지할 때 각 주파수 성분을 선형적으로 인지하지 않고 멜 스케일로 음성을 인지한다. 멜 스케일은 사람이 1kHz 이하에서 민감하게 듣는 것을 이용하여 1kHz 이하 부분은 세밀하게 분석하고, 그 이상은 간격을 넓게 분석하여 청각 구조에 접근시킨 것이다. 이처럼 청각 모델을 기반으로 만들어진 특징을 유해성 오디오의 유해 판별 요소로 사용한다.

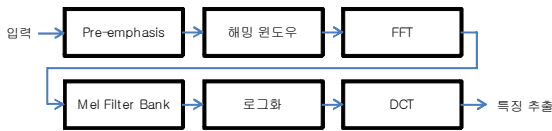


그림 3. MFCC 특징 추출 과정

위의 그림 3은 멜 캡스트럼의 특징 추출 과정이다. 입력된 신호의 프레임에 대해서 pre-emphasis 과정과 해밍 윈도우를 거친 후 FFT를 수행한다. 여기서 FFT 처리는 시계열 데이터의 파형 신호를 주파수 영역으로 변환시키는 과정이다. Mel Filter Bank는 FFT 처리된 결과를 가지고 멜 스케일 필터 값을 곱해주는 과정으로 필터 간 경계 부분의 불연속성을 줄이기 위하여 일반적으로 삼각형 필터를 중첩시켜 사용한다. 로그화 처리는 필터를 통과한 값에 로그를 취하는 것으로 인간의 귀가 소리의 크기에 대해 로그 함수로 느끼기 때문이다. 마지막 DCT 처리는 필터 बैं크의 출력 간의 상관관계를 없애주고 파라미터의 특징을 모아주는 역할을 한다. 특징은 필터 बैं크의 개수만큼 결과가 나오며 본 논문에서는 13차 특징 벡터 계수를 사용하였다.[그림 4]

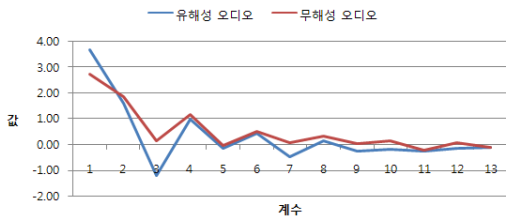


그림 4. MFCC 특징 그래프

3. 유해성 판별

오디오 데이터에서 프레임 단위로 분할하고 특징을 추출하여 유해성을 판별한다. 본 논문에서 프레임 단위는 20ms를 사용하였다. 20ms를 기준으로 특징을 추출하고, 2초만큼 다시 그룹지어 측정하였다. 유해성 오디오는 반복적이고 음이 다채롭게 변하지 않기 때문에 큰 단위인 2초의 그룹으로 분할하고 측정하여 오디오 흐름의 변화에 초점을 두었다. 추출된 특징들의 패턴을 분석하여 각 패턴마다 가질 수 있는 임계치 값을 선정한다. 그리고 영상이 입력되었을 때 특징들을 추출하고 임계치 값에 만족하는지를 측정하여 유해성 판별을 한다.

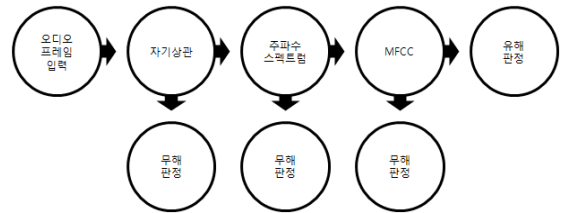


그림 5. 오디오 데이터의 유해 판정 흐름

위의 그림 5와 같이 오디오 데이터의 유해 판정은 자기 상관, 주파수 스펙트럼, MFCC의 순서로 측정을 한다. 순서는 실험을 통해 무해 판별력의 정확성이 높고, 무해 판별력이 큰 순서로 정렬하였다. 처음 정확성과 무해 판별력을 높게 구분함으로써 그 다음 단계로 넘어 갈수록 측정을 하기 위한 비교 시간은 감소된다. 각 측정 단계에서는 해당하는 특징마다 설정한 임계치 값과 비교를 한다. 설정한 임계치 값을 만족하면 유해 상태를 유지하여 다음 단계로 넘어가 비교하고, 만족하지 않는다면 무해 그룹으로 판정한다. 모든 측정 단계에서 임계치 값을 만족하면 해당 그룹은 유해로 판정된다. 동영상 전체 오디오 데이터에서 유해 그룹이 차지하는 비율을 구하고 해당 동영상을 유해 또는 무해로 판별한다.

4. 실험 및 결과

본 논문에서는 제안한 판별 시스템의 성능을 측정하기 위하여 총 10개의 동영상, 오디오 길이는 1분으로 실험을 하였다. 실험 환경은 동영상에서 오디오 데이터만을 가져와 wav로 변환하였으며, 샘플링 주파수는 44100Hz 양자화 레벨은 16비트로 하였다. 구성된 전체 오디오에서 5개는 유해성이 포함된 오디오이고, 나머지 5개는 무해성 오디오이다. 무해성 오디오는 유해가 될 수 없는 음악, 말소리, 기타 외부 환경 음으로 다양하게 구성하였다. 각 오디오 특징마다 추출되는 패턴의 임계치 값을 선정하기 위하여 10개의 오디오 데이터를 사용하였으며 길이는 20초로 하였다. 임계치 값

을 정하기 위한 오디오는 순수하게 유해 또는 무해의 요소만을 담고 있다. 마찬가지로 10개의 오디오를 유해와 무해로 구분하여 구성하였으며 구분된 오디오 데이터를 가지고 특징을 추출한 후 임계치 값을 정하였다. 즉 20초 길이를 가지는 오디오 데이터는 임계치를 설정하기 위한 데이터이고, 1분 길이의 오디오 데이터는 실제 판별에 쓰인 데이터이다.

표 3. 특징에 따른 판별 결과

사용한 특징	유해판별	판별오류	무해판별	판별오류
주파수 스펙트럼	5	1	5	1
자기 상관	8	3	2	0
MFCC	6	2	4	1
ALL	3	0	7	2

표 3은 각 특징에 따른 판별 결과이다. 먼저 주파수 스펙트럼, 자기 상관, MFCC 각각 독립적으로 유해 판별을 하였으며 제안한 다중 특징의 높은 성능을 증명하기 위하여 모든 특징을 조합하여 성능을 측정된 결과와 비교하였다. 성능 측정 평가는 유해 판별의 결과로 유해 판별 오류가 최소인 정확한 판별을 기준으로 정하였으며, 무해 판별의 결과는 오디오 데이터의 유해 판정 흐름 순서에 반영하였다. 독립적으로 특징을 사용하였을 때 보다 모든 특징을 사용한 ALL이 특징별 유해 판정 흐름을 지날수록 유해 판별 수가 적어진다는 점이 있으나, 판별 오류 수가 0으로 가장 정확한 판별을 하였다.

5. 결론

본 논문에서는 주파수 스펙트럼, 자기 상관, MFCC 등 오디오의 특징을 추출하여 유해 동영상 판별을 하였다. 오디오 정보를 이용하여 유해를 판별할 수 있는 요소로 사용할 수 있었으며, 하나의 특징만을 이용하였을 때 보다 다중 특징을 사용하였을 때 판별되는 측정 성능이 높았다. 차후 연구 과제로 판별 오류를 최소화 시키면서 유해 판별의 개수가 적어지지 않는 방향의 연구가 필요하며 제안한 연구와 영상의 이미지 정보를 조합하여 판별하면 더 만족할 수 있는 결과를 얻을 수 있을 것이다.

참고문헌

[1] 유홍식, 김성벽, “청소년의 영상물 이용실태 및 유해 영상물의 영향에 관한 연구”, 사회과학논총, 제13집, Vol.13 pp.95-123, 2006.

[2] 장중수, “내용기반 유해정보 방지기술 개발에 관한 연구”, [IITA] 정보통신연구진흥원 학술기사, 정보통신연구진흥원, 2006.

[3] 김성균, 박명철, 하석운, “동영상의 유해성 판별을 위한

대표 프레임 선정기법”, 한국해양정보통신학회논문지, 1226-6981, 제10권, 제10호, pp.1822-1828, 2006.

[4] 박서영, “UCC 환경에서 Thumbnail을 이용한 유해 동영상 판별 기법”, 석사학위논문 이화여자대학교 교육대학원 : 컴퓨터교육전공, 2008.

[5] 조동욱, “음란콘텐츠에 기반한 유해 음란 사이트의 차단”, 한국통신학회논문지, 제29권, 6B호, 한국통신학회, pp. 554~562, 2004.

[6] I. Mierswa and K. Morik. “Automatic feature extraction for classifying audio data”, Machine Learning Journal, 58:127-149, 2005.

[7] M. Slaney, "A critique of pure audition", Computational Auditory Scene Analysis, 1997.

[8] <http://www.fon.hum.uva.nl/praat/>