

음성 정보를 이용한 성인 콘텐츠 필터링

조정익*, 조진수**, 이일병**

*연세대학교 컴퓨터과학과

**연세대학교 컴퓨터과학과

e-mail:choji@csai.yonsei.ac.kr

Adult Contents Filtering using Speech Information

Jungik Cho*, Jinsu Jo**, Yillbyung Lee**

*Dept. of Computer Science, Yonsei University

**Dept. of Computer Science, Yonsei University

요 약

현재까지 유해한 콘텐츠(Contents)를 차단하기 위한 활발한 연구가 있었으나, 사람의 음성(speech)정보를 이용한 필터링(filtering) 기법에 대한 연구는 활발히 이루어지지 않은 측면이 있다. 본 논문은 동영상 데이터를 가지고 있는 여러 데이터 중에서 음성 정보의 분석을 통하여 일반 콘텐츠와 성인 콘텐츠를 분류하기 위함이다. 본 논문은 음성 정보 중에서 음성 정보의 특징을 가장 잘 다루는 피치 검출을 통한 정보의 분석을 통한 성인 콘텐츠의 필터링에 그 목적이 있다. 현재까지 진행되고 있는 필터링(filtering)방법에 대한 수행 결과보다 개선된 성능을 보이고자 한다. 즉, 음성 정보의 특징 정보를 이용한 성인 콘텐츠(Adult Contents)분류 기법을 활용하는 것으로 성인 콘텐츠(Adult Contents)에서 두드러지는 특징을 보이는 사운드 패턴을 분석하는 것이다.

1. 서론

현재 PC와 네트워킹 기술 등의 급속한 발전으로 말미암아 멀티미디어 데이터의 사용이 많아졌다. 특히, 동영상 데이터는 Web 2.0, 와이브로 등의 발달과 더불어 그 사용이 많이 늘어나고 있다. 이에 편승해 사용자가 만든 여러 멀티미디어 콘텐츠(Multimedia Contents)가 웹 환경에서 쏟아지고 있다. 이런 일련의 변화는 정보의 공유라는 순기능도 있으나, 개인 사생활 정보의 유출, 유해한 정보의 출현 등 부작용도 많이 있다. 본 논문이 제안하는 방법은 동영상 데이터가 가지고 있는 영상 정보, 오디오 정보, 문자 정보들 중에서 동영상에 포함된 음성 정보를 활용하여 유해한 성인 콘텐츠를 필터링하는 기법을 마련하는데 있다. 즉, 오디오 데이터 중에서 사운드 피치 정보를 검출하여 일반 동영상에서 성인 콘텐츠를 필터링하는 방법을 구현하는데 있다. 본 논문의 구성은 성인 콘텐츠에서 오디오를 파형 파일 형식으로 추출하고, COLEA라는 특징 추출 도구를 사용하여 일반 콘텐츠와 성인 콘텐츠의 피치 정보를 추출하였다. 지금까지 널리 알려진 기법보다 보다 향상된 이미지와 사운드 통합 기법을 구현하는데 있다. 본 논문의 일반적인 구성은 2장에서 음성의 특성 분석, 3장은 실험 및 결과에 대해서 설명하고 마지막으로 4장에서 결론을 맺는다.

2. 음성의 특성 분석

음성을 통해서 가장 특징을 잘 분석할 수 있는 것으로 피치(Pitch)정보를 들 수 있다. 즉, 피치(Pitch)정보가 각기 다른 특징을 구분하는데 있어서 가장 중요한 파라미터로

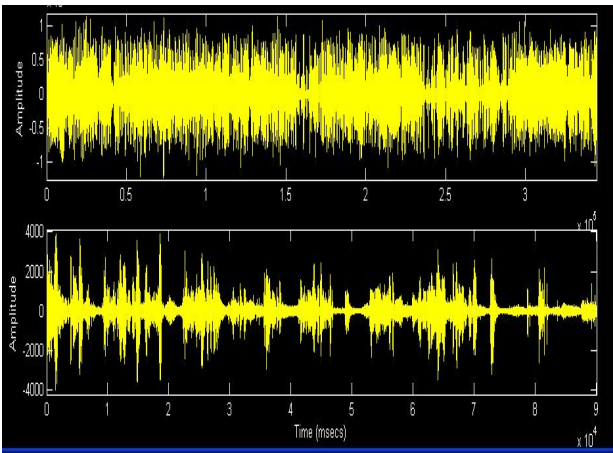
작용함을 알 수 있다. 즉, 음성의 특징을 잘 분류한다고 볼 수 있다. 음성 인식을 위한 피치 정보를 이용한 방법으로는 시간 영역법, 혼성 영역법, 주파수 영역법등이 사용되어 진다. 주파수 영역법은 메모리를 절약하기 위해서 시간 영역법은 파라미터의 정확성을 위해서 주로 사용되어 진다. 이때 사용된 쉘스트림은 DFT나 FFT처럼 신호를 분석하는 알고리즘이다. 쉘스트림은, 푸리에 변환을 log취한 값을 IFT한 값이다. 결과값은 시간 도메인(Time domain)으로 나타난다. 쉘스트림은 음성신호, 지진파측정에 이용되는데, 그 이유는 쉘스트림을 취하면 신호의 포락선이 고유한 패턴을 갖기 때문에 분석하기 쉽기 때문이다. DFT와 FFT가 신호의 주파수 성분을 알아내는 알고리즘이라면 쉘스트림은 DFT나 FFT결과가 크기 값에 대해 명확하지 않는 것을 보완한 알고리즘이다. 따라서 쉘스트림은 지진이나 진동, 음성인식 등 미세 신호 분석에 많이 사용하고 있다. 그리고 스피치(Speech)에서 추출한 데이터를 분석하는 자기상관함수의 푸리에 변환을 한 파워스펙트럼을 사용해 특징을 분석해 보고자 한다. 자기상관함수의 푸리에 변환을 파워스펙트럼이라 하며, 파워스펙트럼의 크기는 각 주파수 성분이 가지는 파워를 나타낸다.

3. 실험 및 결과

동영상 정보의 오디오 영역에서 음성 인식을 위해서 사용한 것은 텍사스 대학교에서 개발한 COLEA이다. 본 논문에서는 다이렉트 쇼(Direct Show)를 통하여 추출된 동영상 데이터의 음성 정보를 파형 파일을 입력 정보로 사용을 하였다. 동영상 콘텐츠가 가진 스피치 데이터를 추출

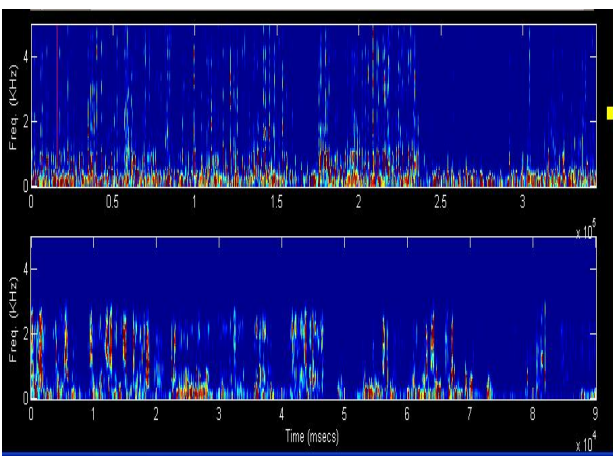
한 후, 스피치는 특정 음원에 대해서 주파수 특징을 추출한 후, 추출한 데이터에서 섀프트립 접근법을 사용하고 특정 주파수에 대한 평균값을 찾아낸다.

피치를 섀프트립 방식에 의해서 평균값을 분석을 해보면 성인 콘텐츠는 130~135Hz의 분포를 보이고 있으며, 일반 콘텐츠는 160~170Hz의 분포를 보이고 있다. 위의 특징 비교 실험 결과를 분석해 보면, 성인 콘텐츠의 경우에는 전체적으로 피치가 낮고, 일반 콘텐츠의 경우에는 전체적으로 피치가 낮다는 것을 알 수 있다. 스피치(Speech)에서 추출한 음원 데이터의 시간에 따른 파형의 형태를 분석하여 일반 콘텐츠와 성인 콘텐츠의 비교를 통하여 특성을 분석한다.



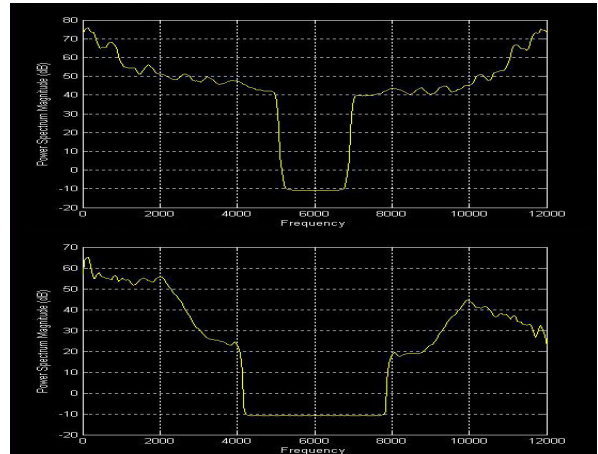
(그림 1) 성인 콘텐츠(위)와 일반 콘텐츠(아래)의 시간에 따른 파형 비교

(그림1)을 살펴보면, 시간에 따른 파형 형태 비교를 통해서 살펴보면 성인 콘텐츠와 일반 콘텐츠가 확연히 다를 수 있다. 위의 차이를 잘 볼 수 있는 방법은 스펙트럼 분포인 (그림2)를 살펴보면 차이를 확연히 알 수 있다. 성인 콘텐츠가 일반 콘텐츠보다 높은 주파수의 빈도가 많은 것을 알 수 있다.



(그림 2) 성인 콘텐츠(위)와 일반 콘텐츠(아래)의 Spectrum 변화비교

이때, 추출한 두 데이터의 파워 스펙트럼 밀도를 분석을 해보면 높은 주파수의 빈도수가 성인 콘텐츠에서 더 많이 보인다는 것을 확인을 할 수 있다.



(그림 3) 성인 콘텐츠(위)와 일반 콘텐츠(아래)의 파워 스펙트럼 밀도 비교

6. 결론

성인 콘텐츠에서 추출된 스피치정보에서 특정 음원정보에 대한 패턴을 구하는 기법을 추가함으로써 현재 많이 연구된 이미지(image)데이터만 활용한 기법보다는 수행처리 능력이 우수할 것으로 판단이 된다. 추후 패턴을 처리하는 방법에서 인공지능 처리 기법을 도입하면 수행 능력 향상에 많은 도움이 될 것으로 생각된다. 본 논문은 동영상 데이터가 가지고 있는 여러 데이터 중에서 음성 정보의 분석을 통하여 일반 콘텐츠와 성인 콘텐츠를 분류하기 위함이다. 음성 정보 중에서 음성 정보의 특징을 가장 잘 다루는 피치 검출을 통한 정보의 분석을 통해서 성인 콘텐츠의 필터링에 그 목적이 있다. 피치 검출을 통한 일반 콘텐츠와 성인 콘텐츠의 특성을 파악함으로써 성인 콘텐츠를 필터링하는 확률을 높일 수 있음을 발견할 수 있다. 본 논문은 동영상에서 나오는 음성에 대한 피치를 검출하고 특성을 분석함으로써 필터링 확률을 높일 수 있다는 계기를 마련했다는 것에서 그 의미를 찾을 수 있다.

참고문헌

- [1] 양진우, 김순협, "성별 구별방법에 의한 자동차 내 음성 인식 성능 향상" 정보과학회논문지: 소프트웨어 및 응용 제 27권 제12호. 2000.12
- [2] 장인창, 박미경, 김태수, 박면웅, "감정변화에 따른 음성정보 분석에 관한 연구" 한국정밀공학회 2004년도 추계 학술대회논문지. 2004
- [3] 조현철, 김우생, "내용기반 검색을 위한 동영상 데이터의 음성인식 모듈 구현" 한국정보과학회 봄 학술발표논문

집 Vol. 26. No. 1. 1999

[4] 김우창, 조선희, 고수정, 이정현, "선별적인 관측열 추출을 통한 DHMM 음성인식의 성능 개선" 한국정보과학회 가을 학술발표논문집 Vol. 27. No. 2. 2000

[5] 전종원, 진용욱, "사상체질과 음성신호의 상관관계" 한국정신과학회 제8회 1998년도 춘계학술대회 논문집. 1998

[6] 김연숙, 김희주, 김경재, "피치 검출과 퍼지화 패턴을 이용한 숫자음 화자 인식에 관한 연구" 한국 컴퓨터정보학회 논문지 제8권 제3호. 2003.9

[7] 최용선(Yong-Sun Choi), 오상훈(Sang-Hoon Oh), 이수영(Soo-Young Lee). "핵심어 검출을 위한 단일 끝점 DTW 알고리즘," 대한전자공학회, 전자공학회논문지-SP 전자공학회논문지 제41권 SP편 제3호, pp. 209-219 (11pages), 2004. 5.

[8] 김종국(Jong-Kuk Kim), 조왕래(Wang-rae Jo), 배명진(Myung-Jin Bae). "음성의 묵음구간 검출을 통한 DTW의 성능개선에 관한 연구," 한국음성과학회, 음성과학 제10권 4호, pp. 117-124 (8pages), 2003. 12.

[9] 우선경, 강철호. "DTW 알고리즘을 이용한 화자적응 신경망에 관한 연구," 대한전자공학회, 대한전자공학회 학술대회 논문집 (신호처리합동) 제8권 1호, pp. 1118-1122 (5pages), 1995. 1.

[10] 정종순, 배재욱, 배명진. "윈도우 환경에서 음성을 이용한 사용자 확인에 관한 연구," 한국음향학회지, Vol.17, No.5. 1998.

[11] 정종순, "대표 평균패턴과 가중 캡스트럼을 이용한 화자인식의 성능 향상에 관한 연구," 석사학위논문, 한국과학기술원, 1996.

[12] 한학용, "MATLAB 실습을 통한 입체적 학습," 패턴인식 개론, 2005.

[13] Furui S. and Sondhi. "Advance in Speech Signal Processing," Dekker.

[14] <http://www.utdallas.edu/~loizou/speech/colea.htm>

[15] <http://blog.naver.com/daban526?Redirect=Log&logNo=120020969762>