

안드로이드 OS 기반 음향 정보를 이용한 유해동영상 검출 서비스의 설계 및 구현

김용운[†], 김봉완^{**}, 최대림^{***}, 고락환^{****}, 김태권^{*****}, 이용주^{*****}

요 약

급속한 인터넷의 발달로 등장하게 된 스마트폰은 여러 가지 긍정적인 모습으로 생활의 편의를 가지고 왔다. 하지만 최근 국내에서 스마트폰의 무분별한 유해물 노출은 사회의 이슈가 되고 있다. 이에 본 논문에서는 안드로이드 OS 기반에서 음향정보를 이용하여 유해동영상을 검출하는 서비스를 설계하고 구현 하였다. 안드로이드 OS기반의 유해동영상 검출 서비스를 구현하기 위해, 기존 음향기반 유해동영상 검출방법의 속도를 향상 시켰다. 검출기로는 GMM(Gaussian Mixture Model)을 사용하였으며, 검출기의 혼합(Mixture)의 수는 18개를 사용하였다. 구현된 서비스의 검출 결과, 일반 동영상 669파일(약 424시간), 유해동영상 541파일(약 263시간), 총 1,210(약 687시간)의 데이터에 대해 97.02%의 검출률로 기존 방법에 비해 검출률은 감소하지 않으면서 속도는 약 5.6배 향상 되었다.

Design and Implementation of Harmful Video Detection Service using Audio Information on Android OS

Yong-Wun Kim[†], Bong-Wan Kim^{**}, Dae-Lim Choi^{***}, Lag-Hwan Ko^{****},
Tae-Guon Kim^{*****}, Yong-Ju Lee^{*****}

ABSTRACT

The smartphone emerged due to the rapid development of the Internet has brought greater convenience to life in a positive manner. Recently, however, because of unconstrained exposure to harmful video, reckless use of smart phones has become a domestic issue in our society. In this paper, a service which detects harmful videos by using the acoustic information is designed and implemented on the Android OS. In order to implement the service of Android OS-based detection of the harmful movie, the speed of existing sound-based detection method for harmful videos is improved. The GMM(Gaussian Mixture Model) was used for classifier and the number of Gaussian Mixture was 18. The implemented service shows a detection rate of 97.02% for a total of 1,210 data files (approximately 687 hours) which comprises 669 general videos files (about 424 hours) and 541 harmful video files (about 263 hours). It's speed is 5.6 times faster than the traditional methods whitout reducing the detection rate.

Key words: Android O.S(안드로이드 O.S), sound based information(음향정보), harmful video(유해 동영상), detection(검출)

※ 교신저자(Corresponding Author): 이용주, 주소: 전라북도 익산시 원광대학교 22 원광대학교 공과대학 5107호 (570-749), 전화: 063)850-7568, FAX: 063)850-7454, E-mail: yjlee@wku.ac.kr

접수일: 2011년 12월 7일, 수정일: 2012년 2월 8일
완료일: 2012년 3월 2일

[†] 정회원, 주식회사 스마트미디어테크
(E-mail: zerokeyw@wku.ac.kr)

^{**} 정회원, 주식회사 보이스웨어
(E-mail: kbw@voiceware.co.kr)

^{***} 정회원, 원광대학교 음성정보기술산업지원센터
(E-mail: dlchoi@wku.ac.kr)

^{****} 정회원, 주식회사 스마트미디어테크
(E-mail: iceblue@wku.ac.kr)

^{*****} 정회원, 원광대학교 음성정보기술산업지원센터
(E-mail: yoplle@wku.ac.kr)

^{*****} 정회원, 원광대학교 컴퓨터공학부

※ 본 연구는 2010학년도 원광대학교 교비 지원에 의해서 연구되었음.

1. 서 론

급속한 무선기기 및 인터넷의 발전은 사람들의 생활 방식을 다양하게 변화시키고 있으며, 최근 등장한 스마트폰은 여러 가지 긍정적인 모습으로 생활의 편의를 제공하고 있으나, 반대로 여러 가지 문제점도 가지고 왔다. 이 중 유해동영상이 무방비로 노출되고 있어 사회의 큰 문제가 되고 있다[1,2]. 이런 유해동영상의 무방비한 노출은 청소년들이 유해동영상에 쉽게 접할 수 있는 환경을 제공한다. 또한 성인의 스마트폰에도 유해동영상을 상업적으로 팔기 위해 제공 업체들이 푸시서비스(Push Service)를 통해 유해동영상을 스마트폰에 보내준다면 사용자가 의도하지 않았더라도 유해동영상을 볼 수 있다.

PC나 서버 환경에서는 유해동영상을 막기 위해 지속적으로 개발이 이루어져 엑스키퍼, 그린아이넷, 아이는 등등 수많은 유해동영상 검출 서비스 프로그램이 존재 한다[3-5]. 그러나 현재까지 스마트폰 환경에서는 유해동영상을 검출하기 위한 서비스가 제공되지 않고 있다.

현재 유해동영상을 막는 대표적인 방법으로는 파일이름, 제목 및 본문 내용 등의 키워드를 이용한 검출 방법과 동영상에 포함된 이미지를 분석하여 유해성 여부를 판단하는 검출방법 등이 있다[6-8]. 그러나 키워드 기반의 필터링 방법은 관련 키워드를 고의로 회피하여 게시하면 이를 검출할 수 없다. 이미지 기반의 필터링 방법은 배경 및 조명상태, 인종에 따른 다양한 피부색 때문에 신체 영역 검출오류로 검출 성능이 저하되는 문제점이 있다.

이러한 유해동영상 검출 기술의 한계를 보완하기 위하여 우리는 키워드 및 이미지 이외에도 중요한 정보를 갖고 있는 음향 신호를 이용하여 유해동영상을 검출하는 기술을 제안한 바 있다[9]. 그러나 이 기술은 PC 또는 서버상에서 이용하기 위해 개발된 것으로, 컴퓨팅 자원이 제한된 스마트폰에서 이용하기 위해서는 검출속도의 문제를 해결해야 한다. 따라서 본 논문에서는 위에서 제안된 기술의 검출 속도를 향상시켜 안드로이드 OS 상에서 유해동영상을 검출하는 서비스를 제안하고 구현한다.

본 논문의 구성은 다음과 같다. 2장에서는 유해동영상에 포함된 음향 신호의 특징에 대해 기술하고, 3장에서는 유해동영상에 포함된 음향 신호의 검출

속도 개선에 대해 기술한다. 4장에서는 안드로이드 OS기반에서 유해동영상 검출 서비스 설계 및 구현에 대해 기술하고, 5장에서는 결론 및 향후 연구를 정리 하였다.

2. 관련연구

2.1 기존 유해동영상 검출 방법 분석

기존 유해동영상 검출 방법에는 크게 키워드 기반 검출방법, 이미지 기반 검출 방법, 오디오 기반 검출 방법이 있다. 키워드 기반 유해 동영상 검출 방법은 국내 UCC 1위 기업인 판도라TV 등에서 1차적으로 유해동영상을 차단하기 위해 사용되는 방법으로 파일이름, 제목 및 본문의 텍스트 내용을 이용하여 검출 검출한다. 유해동영상을 검출 하는 방법 중 가장 초보적이고 1차원적인 필터링 방법이다. 유해동영상을 검출하기 위해 유해물에 관련된 다양한 키워드에 관련된 정보를 데이터베이스에 구축 하고 스트링 매칭 알고리즘을 이용하여 비교 분석한다[10].

하지만 키워드 기반 검출 방법은 키워드를 고의로 회피할 경우 검출 할 수 없다는 단점이 있다. 뿐만 아니라, 유해성에 관련된 키워드가 파일에 존재 하더라도 유해동영상이 아닌 동영상이 많기 때문에, 유해동영상으로 판별 했을 때 2차로 모니터링 작업을 거쳐 직접 사람이 필터링 작업을 마무리해야 된다는 단점도 있다.

이미지 기반 검출 방법은 이미지를 분석하여 유해동영상을 검출 하는 방법이다. 대표적인 이미지 기반 검출 방법으로는 살색 검출 알고리즘을 이용하여 검출하는 방법과 이미지의 패턴정보를 분석하여 검출하는 방법 등이 있다.

살색 검출을 하기위해 사용되는 방법 및 알고리즘은 다음과 같다[11].

- RGB 비율을 이용한 방법

- 대부분의 사람 피부는 멜라닌 색소의 영향으로 인하여 컬러 성분 중 red 성분이 유난히 높은 비율을 차지하고 있음을 이용한 방법으로, R/G, R/B의 비율이 특정 임계치 사이에 존재하면 살색으로 간주하고 그렇지 않으면 살색이 아닌 것으로 검출하는 매우 간단한 알고리즘이다.

- 경험적으로 정의한 살색 영역을 이용하는 방법
 - 몇 가지 색상 공간에서 살색 클러스터들의 경계 점을 살펴보면, 경험적으로 다양한 규칙들을 발견하게 된다. 따라서 이러한 규칙들을 이용한 아래의 수식조건을 만족하면 살색으로 간주하는 알고리즘으로써 매우 간단하고 검출 속도 역시 빠른 관계로 살색 검출이 필요한 연구자들로부터 많이 사용되어진 알고리즘이다.

- Histogram을 이용한 방법

- 살색 픽셀을 찾기 위해 많이 사용되는 방법으로 색상공간을 다수개의 bin으로 양자화한 후, 해당 특정 영역 컬러 성분이 LUT(Look Up Table)로 간주되는 2차원 혹은 3차원 히스토그램을 형성하게 된다. 각 bin들은 training 구간 동안 살색 이미지에서 특정 컬러 성분의 빈도수를 저장하게 되고 그 값이 전체 bin의 합으로 나뉘면 해당 컬러 성분이 살색일 확률을 나타내어 검출하는 방법이다.

- Single Gaussian Model(SGM)을 이용한 방법

- 서로 다른 조명하에서 다양한 사람들의 살색 분포를 살펴보면, 모두 유사한 가우시안 분포를 갖고 있음을 알 수 있는데 이러한 특성을 이용하여 살색 모델을 mean vector와 covariance matrix를 사용하여 2차원 가우시안 모델로 표현한 방법이다.

그러나 살색검출을 이용한 이미지 방법은 특정 영역 등에서 오류가 증가한다는 단점이 있다. 의학 장면의 수술 장면이나, 동영상 화면에 사람의 얼굴이 크게 확대 되어 나온다면 오류가 증가 한다. 뿐만 아니라, 다양한 인종의 피부색으로 인한 오류 또한 증가한다는 단점도 있다.

이미지의 패턴정보를 이용한 방법은 2차원 패턴 정보 사용하여 유해 동영상을 검출 한다. 먼저 유해 동영상의 패턴을 분석하기 위해 데이터를 수집한다. 수집된 데이터를 통해 유해동영상의 주요 특징을 선

택하고, 패턴 인식을 위한 여러 접근법 중에서 어느 모델을 어떠한 알고리즘을 이용하여 어떻게 구성할 것인지를 결정하기 위해 모델을 선택한다. 그 후 수집된 데이터로부터 추출된 특징 집합을 이용하여 유해동영상의 특징들을 학습하고 패턴에 맞게 검출을 하게 된다. 그러나 이미지의 패턴정보를 이용한 방법은 배경 및 조명, 이미지 품질 등에 따라 검출 성능이 저하 된다는 단점이 있다.

오디오 기반 검출 방법은 유해동영상에 포함된 음향정보를 이용하여 검출 하는 방법으로 음성/비 음성 판별, 오디오 유형 분류 및 화자 인식 등에서 자주 사용되고 있는 통계적 음향신호 처리 방법을 이용한 유해동영상 검출 방법이다. 음향 정보를 이용하기 때문에, 배경 및 조명, 피부색 등에 영향을 받지 않는다는 장점이 있으며, 이미지 분석과 비교 하여 상대적으로 단순한 신호분석을 하기 때문에 속도 면에서도 이미지 분석 방법 보다는 스마트폰에 이용하기에 적합하다.

따라서 본 논문에서는 통계적 음향 신호 처리 방법을 이용하여 안드로이드 스마트폰에 유해동영상 검출 서비스를 설계 및 구현 하고자 한다.

2.2 유해동영상에 포함된 음향 신호의 주요 특징

유해동영상에 포함된 음향 신호의 주요 특징으로는 신음소리, 거친 호흡 및 접촉음, 교성 등이 주된 내용을 이룬다. 특히 이러한 소리들은 특정 구간 내에서 일정한 주기를 가지고 반복적으로 나타난다. 아울러 유해동영상의 시나리오에 따른 배우들 간의 음성 대화, 흥미를 끌기 위한 배경음악, 주변 환경소음 등이 포함된 경우가 많다.

음성 신호, 음악 신호 및 유해동영상에 포함된 음향 신호의 특징을 살펴보기 위해 그림 1에 10초 분량의 유해동영상에 포함된 음향 신호, 음악방송의 아나운서 멘트, 가수의 노래가 포함된 록 음악에 대하여 파형과 스펙트로그램을 나타내었다.

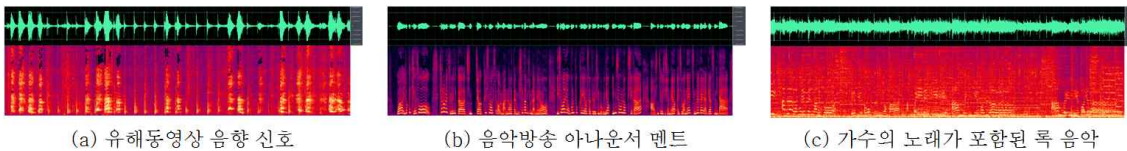


그림 1. 유해동영상 음향 신호, 음성 및 음악 신호의 파형과 스펙트럼

위의 그림 1에서 보는 바와 같이 음성의 경우 자음과 모음의 연속적 발생으로 인해 스펙트럼 포락선의 변화가 록 음악이나 유해동영상의 음향 신호에 비해 빠르다는 것을 알 수 있다. 그리고 록 음악의 경우 비록 빠른 음악임에도 불구하고 스펙트럼 포락선의 변화 속도는 음성에 비해 빠르지 않은 것을 볼 수 있다. 유해동영상 음향 신호의 경우에는 교성과 같은 음향적 특징이 일정한 주기를 가지고 매우 분명하게 반복되어 있음을 볼 수 있다. 유해동영상 음향 신호의 스펙트럼 포락선의 변화 속도가 음성에 비해 느린 것은 신체의 주기적 움직임 속도와 조음 기관의 움직임 속도 차이에 기인한 것이다.

따라서 본 논문에서는 단 구간에서는 음향적 특성을 나타내기 위한 MFCC(Mel-Frequency Cepstral Coefficients)이외에, 주기성 및 변화의 빠르기에 대한 특성을 반영하기 위하여 멜-켄스트럼 모듈레이션 에너지(Mel-Cepstrum Modulation Energy, MCME) 특징[12]을 사용한다. 사람의 청각적 특성을 반영한 MFCC는 멜 스케일로 표현된 주파수 스펙트럼의 형태를 정현파 성분으로 나타낸 것으로 음성인식 및 화자인식 분야에서 가장 널리 사용되고 있다. 그리고 MCME에 대한 정의는 다음과 같다.

$$MCME[n, q] = \frac{\frac{1}{L} \sum_{l=0}^{L-1} \left| \sum_{p=0}^{P-1} C[n+p, l] e^{-j2\pi pq/L} \right|^2}{\frac{1}{P} \sum_{p=0}^{P-1} \log(E[n+p])} \quad (1)$$

여기에서 n 은 프레임 인덱스, l 는 MFCC 계수 인덱스, q 는 모듈레이션 주파수 인덱스, P 는 주파수 분석을 위한 FFT 포인트 수, C 는 MFCC 계수 그리고 E 는 음향 신호의 단 구간 에너지를 의미한다.

3. 유해동영상에 포함된 음향 신호의 검출 속도 개선

기존연구[9]는 PC 또는 서버상에서 이용하기 위

해 개발된 것으로, 특징 추출을 위하여 25ms의 해밍 윈도우를 사용하여 10ms 단위로 프레임을 이동하면서 12차의 MFCC 결과와 단구간 에너지를 추출하였으며, MFCC 추출 시 채널의 영향을 최소화하기 위해 CMN(Cepstral Mean Normalization)이 적용되었다. 1차적으로 추출된 MFCC 결과에 대해 32포인트의 FFT를 프레임 단위로 이동하면서 3.125 Hz ~ 46.88 Hz 범위의 15차의 MCME 를 추출하였다. 그러나 기존연구[9]는 PC또는 서버상에서 이용하기 위해 개발되었기 때문에, 일반적으로 PC보다 하드웨어의 성능이 떨어지는 스마트폰 디바이스 등에서는 계산량을 줄여 검출 속도를 향상시킬 수 있는 방법이 필요하다. 따라서 컴퓨팅 자원이 제한된 스마트폰에서 이용하기 위해서는 검출속도의 문제를 해결해야 한다. 그래서 검출 속도를 개선하는 방법은 아래와 같다.

3.1 검출 속도 개선 방법

유해동영상의 경우 일반적으로 유해성이 있는 장면의 지속 시간이 수분에서 수십 분까지 지속된다는 점을 착안하여 본 논문에서는 검출 속도를 향상시키기 위해 음향 신호에서 일정한 시간 간격을 두고 유해성 검출을 위한 특징을 추출하는 방법을 제안한다.

이를 통해 샘플링 주파수 정규화, MFCC 및 로그 에너지 추출, MCME 특징벡터 추출 등의 신호처리 과정에 필요한 계산량과 시간을 현저하게 절감할 수 있다. 검출 속도를 개선하는 방법은 아래 그림 2에 나타내었다.

특징추출 과정은, 우선 1개의 MCME를 구할 수 있을 길이의 음향 신호를 획득한다. 1개의 MCME를 구할 수 있을 길이(Length)의 계산식은 아래와 같다.

$$Length = P \times A + (W - A) \quad (2)$$

여기에서 P 는 MFCC 특징벡터에서 MCME를 구하기 위한 FFT(Fast Fourier Transform)의 크기,

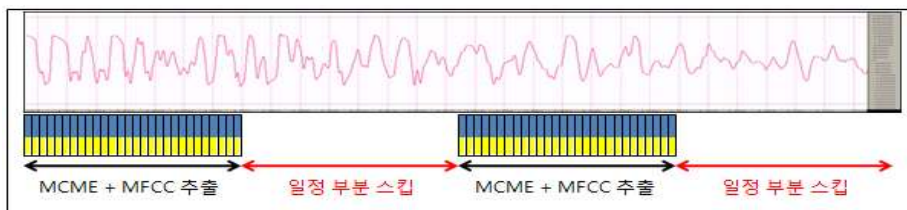


그림 2. 판별 속도를 향상 하는 방법

A 는 MFCC를 구하기 위한 프레임 윈도우의 전진 크기, W 는 MFCC를 구하기 위한 프레임 윈도우의 크기를 의미한다.

1개의 MCME를 구할 수 있을 길이의 음향 신호를 획득하고 샘플링 주파수, 해상도 비트 변환 등의 정규화 과정을 거쳐 P 개의 MFCC와 로그에너지(Log Energy)를 구하고 이로부터 1개의 MCME를 구한다. 추출 과정이 끝난 후 미리 정한 간격만큼 음향 신호를 스킵하고 다시 MFCC 및 MCME을 계산하며, 음향 신호의 끝부분에 도달할 때까지 이러한 과정을 반복한다.

3.2 실험 및 평가

3.2.1 검출 실험을 위한 데이터베이스 구성

유해음향 모델 및 안티 모델의 학습, 개발, 평가를 하기 위해 표 1과 같이 음향 DB를 구성 하였다. 학습 데이터의 경우 기존 연구[9]와 동일한 데이터를 사용하였으며, 나머지 데이터 중 각 유형별로 임의로 20%의 데이터를 선정하여 개발 데이터로, 나머지 80%를 평가 데이터로 사용 하였다.

음향 기반 판별을 위한 전처리 과정으로서 모든 동영상 및 음향 데이터로부터 11kHz, 16bit, Mono 포맷으로 음향 신호를 추출하였다.

3.2.2 실험 환경 및 고속처리를 위한 스킵길이의 결정

학습, 개발, 평가를 위해 유해동영상의 음향 데이

터에서 유해성이 나타나는 부분만을 세그멘테이션하는 작업은 전혀 수행하지 않았으며, 전체 파일을 사용하였다. 특징 추출을 위해 25ms의 해밍윈도우를 사용하여 10ms 단위로 프레임을 이동하면서 12차의 MFCC결과와 단구간 에너지를 추출 하였으며, MFCC 추출 시 채널의 영향을 최소화하기 위해 CMN (Cepstral Mean Normalization)이 적용되었다. 1차적으로 추출된 MFCC결과에 대해 32포인트의 FFT를 프레임 단위로 이동하면서 15차의 MCME를 추출 하였다.

실험에 사용된 검출기는 GMM(Gaussian Mixture Models)을 사용하였으며, GMM 혼합(Mixture)의 수는 18개를 사용하였다. 학습 데이터로부터 유해한 오디오 신호를 대표하는 유해 모델 λ_o 과 일반적인 오디오 신호 중 음악을 대표하는 모델 λ_M 그리고 기타 정상 오디오 신호를 대표하는 모델 λ_G 를 구성하여 학습하였다. 최종 판단은 다음 식 3과 같이 우도비(likelihood ratio)를 이용하여 판단하였다.

$$C = \frac{p(X|\lambda_o)}{p(X|\lambda_o) + \max(p(X|\lambda_M), p(X|\lambda_G))} \quad (3)$$

여기에서 X 는 동영상에서 추출한 특징 벡터의 열을 의미하며, $p(X|\lambda_i)$ 는 GMM 모델 i 에 특징 X 가 나타날 확률이다. 따라서 분자(numerator)는 동영상에서 추출한 특징 벡터를 이용하여 계산한 유해일 확률을 의미하며, 분모(denominator)는 전체 사운드에 대한 백그라운드 모델(background model)의 확률을

표 1. 데이터베이스의 구성, 학습, 개발 및 평가세트

구분	유형	학습용		개발용		평가용		계	
		파일수	시간	파일수	시간	파일수	시간	파일수	시간
일 반	음악	87	7.0	55	3.7	222	21.0	364	31.7
	뉴스	5	4.2	16	12.3	64	50.2	85	66.7
	다큐멘터리	5	3.8	18	13.7	74	57.1	97	74.6
	예능	0	0	12	13.9	48	57.9	60	71.6
	영화	6	7.1	15	19.8	64	80.7	85	107.6
	드라마	0	0.0	14	14.7	58	62.0	72	76.7
	애니메이션	0	0.0	16	5.6	62	16.0	78	21.6
	스포츠	2	0.8	10	8.2	38	36.0	50	45.0
	인터넷 강의	0	0.0	10	11.2	39	43.2	48	54.4
계	105	22.9	166	103.1	669	424.1	940	550.1	
유해	인터넷에서 구한 유해동영상	30	7.9	135	56.6	541	262.4	706	326.9
총계		135	30.8	301	159.7	1210	686.5	1646	877.0

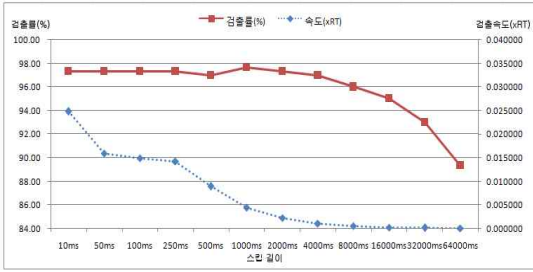


그림 4. 개발 세트의 검출률 및 검출 소요시간

의미하게 된다. 위와 같이 계산된 C 값은 항상 0에서 1사이의 값을 갖게 되며 0.5 이상일 경우 유해 사운드의 확률이 높음을 의미한다. 본 논문에서는 계산된 C가 0.5이상일 경우 유해로 판단하였다.

스킵 길이를 결정하기 위해서 스킵 길이를 다양하게 변경하면서 실험하였다. 음향 신호에서 특징 추출을 위한 스킵 길이가 클 경우 이에 비례하여 검출 속도가 증가하지만, 검출률은 저하될 가능성이 있다. 따라서 임의의 동영상에 입력되었을 때 검출률의 저하는 최소화 하면서 속도 향상을 이룰 수 있는 최적의 스킵 길이를 확인하기 위하여 표 1의 개발 세트를 이용하여 실험 하였다. 그 결과는 그림 4와 같다. 여기에서 검출속도(xRT)는 검출에 소요되는 시간을 데이터 시간으로 나눈 것으로 작을수록 빠르다. 속도(xRT) 측정은 Intel® core™2 Quad Q9650 @ 3.0GHz, 4GByte의 RAM을 갖는 Windows XP 환경에서 수행 하였으며 동영상 데이터에서 음향 데이터를 추출하는 시간은 포함되지 않았다.

표 3. 검출 결과

구분	평가 데이터		검출 결과		
	유형	파일수	일반	유해	검출률(%)
일	음악	222	220	0	100.00
	뉴스	64	64	0	100.00
	다큐멘터리	74	74	0	100.00
	예능	48	48	0	100.00
	영화	64	61	1	98.44
반	드라마	58	58	0	100.00
	애니메이션	62	58	0	100.00
	스포츠	38	32	8	78.95
	인터넷 강의	39	39	0	100.00
	계	669	654	9	98.65
유해	인터넷에서 구한 유해동영상	541	27	514	95.01

표 2. 기존 방법과 제안된 방법의 검출률 및 속도 비교

구분	검출률(%)	속도(xRT)
기존 방법[9]	97.02	0.02044
제안된 방법	97.02	0.00366

그림 4에 나타난 바와 같이 500ms에서 속도가 급격하게 증가를 하는 것을 볼 수 있으며, 1,000ms 이후부터는 검출률이 점차적으로 감소를 하고 있다. 따라서 본격적인 검출 성능평가에서는 스킵 길이를 1,000ms으로 고정하여 실험을 수행 하였다.

3.2.3 검출 성능평가

기존 방법과의 성능 비교를 위해 기존에 제안된 10ms의 방법과 개발세트를 이용하여 결정된 1,000ms의 방법에 대해 검출 및 속도 실험을 하였다. 총 1,210개(약687시간)의 파일로 이루어진 평가 세트를 이용 하였으며, 이에 대한 총 검출률과 속도비교를 표 2에 정리 하였다.

평가 세트를 이용한 실험에서는 표 2에 보는 바와 같이 기존 방법과 비교하여 검출률은 동일하였으나 속도는 약 5.6배 정도 향상되어 본 논문에서 제안하는 속도 향상 방법이 유효함을 알 수 있다. 검출률에 대한 자세한 분석을 하기위해 제안된 방법의 검출 결과를 표 3에 작성 하였다.

일반 동영상의 경우 스포츠 데이터에서 8개의 검출 오류를 보였으며, 영화 데이터에서는 1개의 검출 오류를 보였다. 스포츠 데이터의 경우 선수들의 호흡

음 충격음 등이 자주 발생하여 그 음향적 특성이 음성이 주도하는 일반데이터와 다소 다른 점을 고려하면, 별도의 유형으로 모델링 할 경우 검출 성능을 향상시킬 수 있으리라 기대한다. 영화 데이터에서는 영화 내용 중 유해성 있는 부분이 일부 포함 되어 있어 유해동영상으로 검출 되었다.

유해동영상 오류의 경우 영상으로는 유해 부분이 많이 포함되어 있지만 27개 모두 다 강한 배경음을 가지고 있어 유해성이 있는 음향정보가 거의 드러나지 않는 데이터였다.

4. 안드로이드 OS 기반에서 유해동영상 검출 서비스 설계 및 구현

4.1 설계

본 논문에서 설계한 서비스에 대한 전체적인 흐름은 그림 5와 같다. 왼쪽의 그림은 안드로이드 OS 기반 음향정보를 이용한 유해동영상 검출 서비스 흐름도이고, 오른쪽의 그림은 그에 따른 순서도이다.

안드로이드 스마트폰은 동영상 파일이 스마트폰 SD메모리에 저장될 때, 자동으로 미디어스캔 이벤트(Media Scan Event)가 발생 하게 된다. 미디어스캔 이벤트가 완료되면 안드로이드 미디어 데이터베이스에서 저장된 파일의 위치 정보를 얻어와 기존에 검색하였던 동영상 파일인지를 1차적으로 우선 비교하고, 기존에 검사가 이루어지지 않은 파일이 존재

할 경우에만 검출작업을 진행한다. 검출 작업이 진행되면 동영상에서 음향 정보를 구하고 특징 벡터를 추출한다. 추출된 특징벡터를 GMM 검출기를 이용하여 유해성 여부를 검출한 후 검사이력 등에 관련된 DB를 수정하고 유해동영상 검출작업을 완료하였다. 이는 브로드캐스트(Broadcast)를 전송한다. 그 후 안드로이드 미디어스캔 이벤트가 발생할 때 까지 대기 상태로 돌아간다. 안드로이드 미디어스캔 이벤트가 언제 발생 할지 모르기 때문에 스마트폰이 켜짐과 동시에 서비스도 같이 실행되어 대기할 수 있게 설계 하였다.

안드로이드 OS기반 음향정보를 이용한 유해동영상 검출 서비스는 그림 6의 (a)와 같이 크게 Media Scanner, Android Media Defender Service, Android Native Libs, Android File System으로 구성되며, 그림 6의 (b)에 데이터 및 메시지 흐름도를 나타내었다.

서비스에서 어플리케이션과 상호 연동을 하기 위해 자바 인텐트(Intent) 객체를 이용하였다. 인텐트 식별자는 표 4와 같으며, 검출 결과에 대한 메시지 구조는 그림 7과 같다.

4.2 구현 및 성능평가

안드로이드 OS에 구현된 서비스 화면은 그림 8 (a)에 표현 하였다. 구현된 서비스를 평가하기 위해 페어랜탈 목적의 테스트 어플리케이션을 구현하였다. 테스트를 위한 어플리케이션의 구성화면은 그림

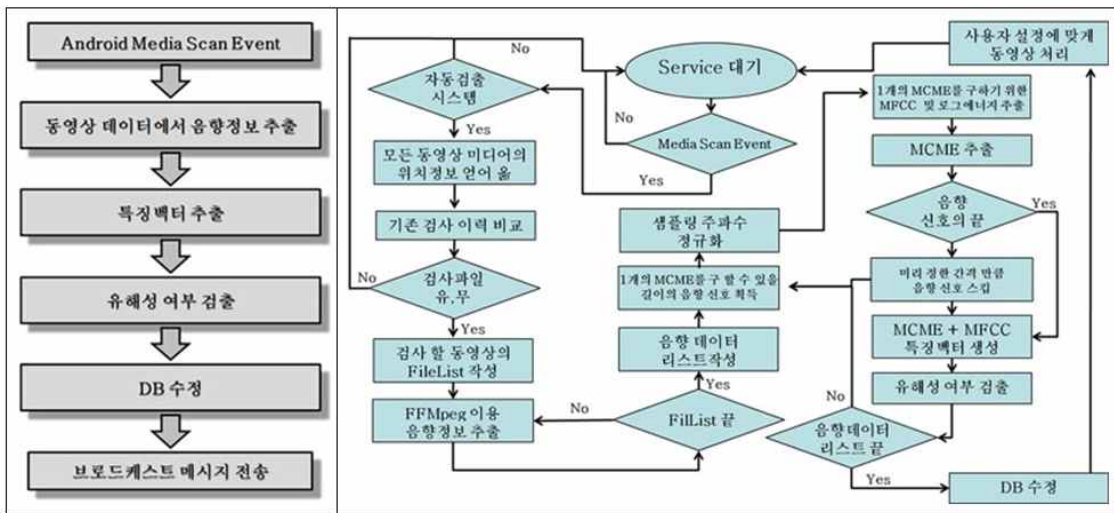


그림 5. 음향정보를 이용한 유해동영상 검출 서비스 흐름도 및 순서도

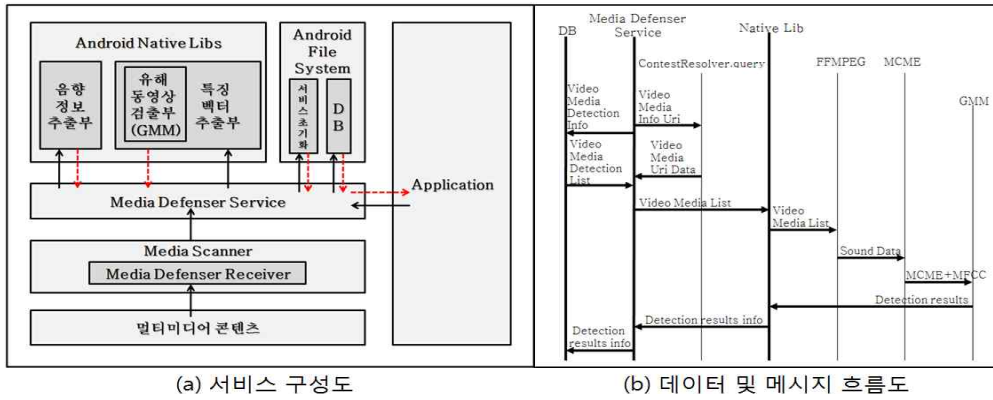


그림 6. 음향정보를 이용한 유해동영상 검출 서비스 구성도, 데이터 및 메시지 흐름도

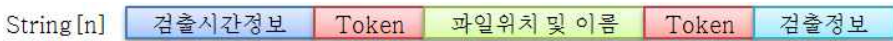


그림 7. 요청한 DB정보의 메시지 구조

표 4. 어플리케이션과 서비스간의 인텐트 식별자

Action	System		Application		Media Defender Service	
	Send	Receiver	Send	Receiver	Send	Receiver
검출정보 요청 및 완료	Action_DB_Info	-	-	Action_DB_Send	Action_DB_Send	Action_DB_Info



그림 8. 구현된 서비스 및 테스트 어플리케이션 화면

8의 (b)~(d)와 같다.

구현된 서비스와 테스트 어플리케이션은 삼성 갤럭시S에 설치하여 성능 평가를 하였다. 구현된 스마트폰의 환경사항은 아래 표 5와 같다.

표 5. 스마트폰 구현 및 실험환경

CPU	1GHz
RAM	326Mb
기기명	삼성 갤럭시 S
OS	Android 2.3

스마트폰에서는 하드웨어의 특성상 PC처럼 대용량의 데이터를 저장해 놓을 공간이 없으므로 구현된 서비스의 성능분석을 위해 1시간 20분 분량의 데이터를 이용하여 속도 측정을 하였다. 기존 방법에서는 0.05865(xRT), 제안된 방법에서는 0.01047(xRT)의 검출 속도가 측정되었으며 이는 기존 방법의 경우 1시간의 동영상을 처리할 때 약 3분 30초, 제안된 방법의 경우 약 38초의 시간이 걸린다는 것을 나타낸다.

5. 결론 및 향후 연구

본 논문에서는 기존의 MCME+MFCC 특징벡터를 이용하여 유해동영상 검출 방법의 속도향상 방법을 제안 하였으며, 안드로이드 OS 기반에서 음향정보를 이용한 유해동영상 검출 서비스를 설계하고 구현하였다. 기존 방법에 비해 속도는 약 5.6배 향상되었다. 뿐만 아니라, 검출률도 97.02%로 기존연구와 동일하여 유효함을 증명하였다. 본 논문에서 제안한 속도 향상 방법을 스마트폰이 아닌 PC나 서버에서도 이용한다면, 같은 시간에 더 많은 유해동영상을 검출할 수 있을 것이다.

구현된 서비스는 다른 어플리케이션에서 쉽게 접근하고 활용할 수 있도록 하였기 때문에 어플리케이션 개발자는 유해동영상 검출에 필요한 사전 지식이 없어도 본 논문에서 구현된 서비스를 이용할 수 있다.

향후 연구 방향으로는 추가적 속도 향상을 위한 방법 및 스포츠 영역등 오류 유형에 대한 검출 성능 향상 방법에 대한 연구가 필요하다.

참 고 문 헌

- [1] 정보통신연구진흥원, 주간기술동향 통권 1332호, 2008.
- [2] 박영호, 임선영 “차세대 안드로이드 기술동향 분석 및 앱 개발 사례,” 멀티미디어학회논문지, 제15권, 제1호, pp. 74-84, 2011.
- [3] 엑스키퍼, <http://www.xkeeper.com/>, 2012.
- [4] 그린아이넷, <http://www.greeninet.or.kr/>, 2012.
- [5] 아이눈, <http://www.aiyac.com/>, 2012
- [6] M. Hammani, Y. Chahir, and L. Chen, “Web-Guard : A Web Filtering Engine Combining Textural, Structural, and Visual Content-Based Analysis,” *IEEE Trans. on Knowledge and Data Engineering*, Vol.18, No.2, pp. 272-284. 2006.
- [7] H. Lee, S. Lee, and T. Nam, “Implementation of High Performance Objectionable Video Classification System,” *Proc. ICACT2006*, pp. 959-962, 2006.
- [8] W. Kim, H. Lee, J. Park, and K. Yoon, “Multi Class Adult Image Classification using Neural Networks,” *LNAI*, Vol.3501, pp. 222-226, 2005.
- [9] 김봉완, 최대림, 이용주 “오디오신호에 기반한 음란 동영상 판별,” 대한음성학회 논문지, 제63호, pp. 139-152, 2007.
- [10] 이티뉴스 (2012. 12. 5), http://www.etnews.com/list/press_view.html?id=0235166
- [11] 김광훈, 권준찬, 송우진, “살색검출을 기반으로 한 포르노 영상 필터링,” 제 16회 신호처리 합동 학술대회 논문집, pp. 77-80, 2003.
- [12] B. Kim, D. Choi, and Y. Lee, “Speech/music Discrimination using Mel-Cepstrum Modulation energy,” *LNAI*, Vol.4629, pp. 106-414, 2007.



김 용 운

2009년 2월 원광대학교 컴퓨터공학부 공학사
2011년 2월 원광대학교 컴퓨터공학부 공학석사
2011년 2월 ~ 현재 주식회사 스마트미디어테크 연구원

관심분야 : 음성정보처리, 오디오 분류, 멀티미디어응용



고 락 환

2002년 2월 원광대학교 컴퓨터공학과 공학사
2006년 2월 원광대학교 컴퓨터공학과 공학석사
2006년 3월 ~ 2011년 2월 원광대학교 음성정보기술산업지원센터 연구원

2011년 3월~현재 주식회사 스마트미디어테크 대표이사
관심분야 : 음성정보처리, 멀티미디어응용, HCI



김 봉 완

1995년 2월 원광대학교 컴퓨터공학과 공학사
1997년 2월 원광대학교 컴퓨터공학과 공학석사
2002년 2월 원광대학교 컴퓨터공학과 공학박사

2001년 2월~2012년 2월 원광대학교 음성정보기술산업지원센터 연구기획실장
2012년 3월~현재 주식회사 보이스웨어 차장
관심분야 : 음성인식, 오디오 분류



김 태 권

2009년 2월 원광대학교 컴퓨터공학부 공학사
2011년 2월 원광대학교 컴퓨터공학부 공학석사
2011년 2월~원광대학교 음성정보기술산업지원센터 연구원

관심분야 : 음성정보처리, 오디오 분류, 멀티미디어응용



최 대 림

2000년 2월 전북대학교 전자공학과 공학사
2002년 2월 전북대학교 전자공학과 공학석사
2009년 8월 전북대학교 전자공학과 공학박사

2002년 5월~현재 원광대학교 음성정보기술산업지원센터 책임연구원
관심분야 : 음성인식, 음성합성



이 용 주

1976년 2월 고려대학교 전자공학과 공학사
1987년 8월 고려대학교 전자공학과 공학석사
1992년 8월 고려대학교 전자공학과 공학박사

1980년 8월~1994년 2월 학국전자통신연구원 책임연구원
1994년 3월~현재 원광대학교 컴퓨터공학부 정교수
관심분야 : 음성정보처리, 멀티미디어응용, HCI